Discovering and Engineering Novel Thermostable Terpene Synthases

Nesbitt, Edward

Award date:
2020

Awarding institution:
University of Bath

Discovering and Engineering Novel Thermostable Terpene Synthases

Edward Andrew Nesbitt

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2019

Abstract

Terpenes are the most diverse group of natural products with more than 80000 structures and have a wide range of applications in industry ranging from pharmaceuticals to flavours and fragrances. Most commercial terpene are extracted from plants, but this method typically results in low yields. Microbial platforms such as the *Saccharomyces cerevisiae* and *Rhodobacter sphaeroides* platforms can provide a cheaper, more sustainable alternative. Recently, the first thermostable terpene platform was developed in the thermophile, *Parageobacillus thermoglucosidasius*. As well as consuming the breakdown products of waste lignocellulosic biomass as a feedstock rather than sugar like the other microbial terpene platforms, this platform is believed to have other advantages including lower risk of contamination, increased substrate solubility and lower running costs due to running at high temperatures. One of the final steps in the terpene pathways, catalysed by the terpene synthase (TPS), is mainly responsible for the structural diversity of terpenes. Before this work, the *Parageobacillus* platform could only synthesise the sesquiterpene, τ-muurolol, as no other thermostable TPSs had been identified other than the two τ-muurolol synthases from *Roseiflexus* species. For this system to be industrially viable, valuable terpenes need to be produced in high yields.

To increase the number of terpenes produced by the *Parageobacillus* platform, this work aimed to characterise more thermostable TPSs. Hidden Markov Models (HMMs) were used to search and identify novel thermostable TPSs from thermophiles which were then characterised using *in vitro* assays. This strategy identified the first naturally thermostable germacrene D-4-ol synthase as well as the first (+)-sativene synthase. In order to identify methods of increasing the thermostability and robustness of mesostable TPSs without screening large numbers of mutants, structural comparisons between mesostable and thermostable τ-muurolol synthases were conducted. Computational methods were then used to identify thermostabilising mutations in mesostable τ-muurolol and selinadiene synthases which resulted in small increases in thermostability. Finally, the active site of the thermostable τ-muurolol synthase, RoseRS_3509, was mutated to create a TPS that synthesised the prospective biofuel, β-farnesene. This work also led to increased understanding about which active site amino acids are involved in the cyclisation mechanism.

Acknowledgements

I would like to thank my supervisor, Professor David Leak, for the opportunity to work on this project and in Lab 1.28. You gave me the freedom to make this project my own from the outset but gave me support when it was required as well as pushing me to take on my fear of presenting. I would also like to thank my co-supervisor, Dr Susan Crennell, for the regular support and for always being available at short notice to talk to about experiments. I express my gratitude to BBSRC for funding my project.

I am so thankful to everyone in the Leak and Pudney labs, past and present, who have helped me from when I was a shy, anxious final year undergraduate project student to what I have become today. I am extremely appreciative of all the help from Dr Styles and Dr Chacon with all of my terpene work as well as the undergraduate and master's students I have supervised. I would also like to thank Rory, Hannah, Dragana and Alex, for their much-needed help with everything protein related. To Emanuele, Rory and Helen, I will miss our spontaneous Chinese dinners on our walk home. To Mavi, it has been my pleasure to work alongside you from our first day of sharing a lab bench to our graduation together. To Liz, you have been a great friend throughout our PhDs and I can't thank you enough for all of the support throughout. I hope I provided just as much support in return. To Abby, thank you for all the time you put into helping me find strategies to deal with my anxiety. All of you have made my PhD immensely enjoyable and provided many great memories that I will take away with me.

To my parents, this is entirely your fault. For my whole life, you made science unavoidable and piqued my interest at a young age. Since then you have supported me every step of the way and given me a wealth of opportunities to help me get this far. To my brother, Haza, you continue to motivate me to be a better person even though we still bicker in our mid-20s.

Lastly to my partner in crime, Rebecca. I couldn't have done this without you. We have been on a rollercoaster of a journey over the past six years where you have supported me through the darkest times and celebrated with me at the best of times. You gave me the confidence me to take opportunities and try new things that I would not have been able to do on my own. Thank you for making me a better person. You are incredible, you are my champion and I am truly "greatful" that you were the one I got to share the PhD experience with. I am excited to find out what the future holds for us.

Contents

Figures List

Tables List

Abbreviations

µg	Microgram
µl	Microlitre
°C	Degrees Celsius
aa	Amino acids
AACT	Acetoacetyl-CoA Thiolase
Amp	Ampicillin
βME	β-mercaptoethanol
bp	Base pairs
BSA	Bovine Serum Albumin
CD	Circular dichroism
Chl	Chloramphenicol
CoA	Coenzyme A
cPCR	Colony polymerase chain reaction
dH_2O	Deionised water
DMAPP	Dimethylallyl Diphosphate
DMD	Diphosphomevalonate decarboxylase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FID	Flame ionisation detector
FPP	Farnesyl diphosphate
FPPS	Farnesyl Diphosphate Synthase
g	Gram
GC-FID	Gas chromatography flame ionisation detector
GC-MS	Gas chromatography mass spectrometry
GPP	Geranyl diphosphate

GGPP	Geranylgeranyl diphosphate
His-tag	Hexahistidine tag
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HMGR	3-hydroxy-3-methylglutaryl-CoA Reductase
HMGS	3-hydroxy-3-methylglutaryl-CoA Synthase
hpi	Hours post-induction
hr(s)	Hours
IDI	Isopentenyl Diphosphate Synthase
IPP	Isopentenyl Diphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
Kan	Kanamycin
kb	Kilobase
kDa	Kilodalton
LB	Luria Bertani Broth
L	Litre
Mbp	Mega base pairs
MCS	Multiple cloning site
MEP	2-methyl-D-erythritol-4-phosphate
mg	Milligram
min(s)	Minute(s)
ml	Millilitre
MK	Mevalonate Kinase
MVA	Mevalonate
MWCO	Molecular weight cut-off
NIST	National Institute of Standards and Technology
nm	Nanometre
NPP	Nerolidyl diphosphate
nt	Nucleotide
OD_x	Optical Density at x nm

PCR	Polymerase chain reaction
pH	Potential hydrogen
PMK	Phosphomevalonate Kinase
rpm	Rotations per minute
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
secs	Seconds
SOC	Super Optimal Broth with Catabolite
T_m	Melting temperature
TPS	Terpene synthase
TAE	Tris acetate EDTA
w/v	Weight/volume
WT	Wild type
XIC	Extracted ion content

1 Introduction

1.1 Natural products

The chemicals that are produced by all living organisms in the three domains of life are known as natural products. Many of these chemicals are vital to the survival of an organism and these are known as primary metabolites. The primary metabolism is responsible for the biosynthesis of key chemicals including amino acids, lipids, nucleic acids and carbohydrates that can be used as energy resources or individual units for macromolecules including proteins, polysaccharides and DNA/RNA. All of these key chemicals involved in primary metabolism need to be maintained at precise levels for cells to remain viable otherwise the organism would not be able to grow, develop and reproduce. The pathways responsible for primary metabolism are highly conserved among related species. In contrast, the chemicals that are not essential for cell survival are known as secondary metabolites. Secondary metabolites have an incredibly diverse range of structures but they are considered metabolically costly as they are not vital and divert valuable resources away from primary metabolism. Due to the expense of these chemicals, secondary metabolites are only made under specific conditions (Tyc *et al.* 2016, Yang *et al.* 2018). Although costly, the production of a secondary metabolite provides a significant advantage as it enables interaction or competition between neighbouring organisms. For example, bacteria can release compounds such as antibiotics that can inhibit other microorganisms in the competition for space and nutrients or siderophores to eliminate or scavenge iron and deprive competitors (Hider and Kong 2010, Sultan *et al.* 2018). Many plants release a range of volatile compounds to repel herbivores and/or lure herbivore predators as forms of protection (Hare 2011, McCormick *et al.* 2012).

Throughout history, humans have used natural products from both primary and secondary metabolism in traditional medicines, essential oils and alcoholic drinks (Buckley *et al.* 2004, Kirby and Keasling 2008). Nowadays, natural products from the primary and secondary metabolism of various organisms are used in industries including pharmaceuticals, flavour and fragrance, food and drink, cosmetic and biofuel. They can be produced and isolated through microbial fermentation, plant extractions, enzymatic processes or chemical synthesis (Sun *et al.* 2015, D'este *et al.* 2018, Sanchez *et al.* 2018, Acevedo-Rocha *et al.* 2019). The latter technique is unfavourable for many reasons including the production of toxic side products or

waste, costly processes and the inability to control the stereospecificity of the reactions. The other three techniques provide cleaner, renewable alternatives (Sun *et al.* 2015, D'este *et al.* 2018).

Several key industrially produced primary metabolites include alcohols, amino acids, organic acids and vitamins (Singh, Kumar, *et al.* 2017). Amino acids are vital for protein biosynthesis in every organism. Commercially, amino acids can be used for a wide range of applications, such as supplements for dietary needs or treatments for various disorders and diseases as well as in food as flavour enhancers or antioxidants resulting in the amino acid market being very large (Singh, Kumar, *et al.* 2017, Sanchez *et al.* 2018, Liang *et al.* 2019). Most of the amino acids, such as L-tryptophan, L-cysteine or L-glutamate, are available commercially through microbial fermentation, where engineered *E. coli* and *Corynebacterium glutamicum* are mainly used for production (Sun *et al.* 2015, Liang *et al.* 2019). However, L-methionine still requires chemical synthesis (Sanchez *et al.* 2018). In nature, animals source vitamins from bacteria and plants which naturally produce them. Many vitamins are precursors to coenzymes or cofactors that are vital for the function of various enzymes (Acevedo-Rocha *et al.* 2019). Vitamins are therefore used as dietary supplements and in animal feed for health. Many commercially available vitamins, such as riboflavin and γ-linoleic acid, are produced by fermentation by various bacteria or fungi. However, many commercial vitamins are still chemically synthesised so that they can be including biotin, niacin and ascorbic acid (Demain 2007, Singh, Kumar, *et al.* 2017, Acevedo-Rocha *et al.* 2019). Organic acids, such as citric acid or lactic acid, are used in a range of industries from food and drink to chemical feedstocks. For example, lactic acid can be used as a preservative in food while its polymer, polylactic acid, can be used as a biodegradable plastic (Martinez *et al.* 2013, Sun *et al.* 2015, Juturu and Wu 2016). Most organic acids can be produced commercially by microbial fermentations using various yeast, bacterial and fungal species (Sauer *et al.* 2008, Klement and Büchs 2012, Jansen and van Gulik 2014, Show *et al.* 2015). Lastly, many organisms produce alcohols as the final product of various pathways during fermentation including ethanol and butanol. For commercial ethanol production, engineered yeast strains are mainly used. However, the commercial production of butanol still uses chemical synthesis as well as Clostridial bacteria (Demain 2007, Ndaba *et al.* 2015, Azhar *et al.* 2017). The leading use of alcohols is as biofuels but they are also used as chemical feedstocks and solvents. Bioethanol is the current leading renewable alternative to petroleum-based fuels (Azhar *et al.* 2017).

There are multiple classes of natural products from secondary metabolism. The major secondary metabolite classes from plants are alkaloids, phenolic compounds, polyketides (PKs) and Nonribosomal peptides (NRPs) and terpenes (Flores-Sanchez and Verpoorte 2009, Chomel *et al.* 2016). In addition, bacteria and fungi also produce non-ribosomal peptides (NRPs). There is some crossover between the classes as hybrid compounds can be formed such as the anticancer drug, taxol, which is a plant terpene alkaloid (Demain and Vaishnav 2010). The majority of secondary metabolites have been isolated from plants with a fraction being identified as having useful industrial applications. However, for many of these compounds, due to the complexity of the structures and the pathways required to synthesise them, industrial production still relies on extraction from the native organism, where the desired compound is only present in low yields. Only a few compounds have had sustainable microbial platforms developed to date.

Alkaloids are characterised as cyclic structures that contain at least one nitrogen in one or more of the cycles (Wang, Guleria, *et al.* 2016, Singh, Kumar, *et al.* 2017). Currently, over 18000 alkaloid structures have been isolated, where the majority are produced by plants and rest are made by bacteria, fungi and animals (Cushnie *et al.* 2014). They are often synthesised from amino acids such as L-aspartate, L-glutamate and the aromatic amino acids although they can also be synthesised by other routes (Schäfer and Wink 2009, Kallscheuer *et al.* 2019, Qin *et al.* 2019). Alkaloids are believed to have a range of natural functions from being a predator deterrent, to an allelochemical or antimicrobial (Cushnie *et al.* 2014). Alkaloids are of particular interest to the pharmaceutical industry as their applications range from anti-cancer and antimalarial to pain relief and blood pressure control. However, no microbial platforms are commercially available for alkaloids and chemical synthesis is often not a cost effective option for the complex alkaloid structures (Nakagawa *et al.* 2011). The only commercial option is extraction from the native plant which produces low yields (Nakagawa *et al.* 2011, Diamond and Desgagné-Penix 2016). An example is galanthamine which is commercially produced by extraction from species of the *Amaryllidaceae* family for the treatment of Alzheimer's disease (Heinrich and Teoh 2004, Halpin *et al.* 2010, Fraser *et al.* 2017). To try and produce increased yields of alkaloids, many studies are looking at using microbes for the production of these complex molecules (Nakagawa *et al.* 2011, Diamond and Desgagné-Penix 2016, Wang, Guleria, *et al.* 2016).

Phenolic compounds include any secondary metabolites synthesised from intermediates of the shikimate pathway such as benzoates, tannins and flavonoids

(Chomel *et al.* 2016, Kallscheuer *et al.* 2019). Many phenolic compounds are used as signalling molecules by plants or have antimicrobial activity (Weston and Mathesius 2013, Mierziak *et al.* 2014). Many of these compounds have industrial applications such as pharmaceuticals, pesticides and flavours and fragrances. The isolation of individual phenolic compounds by plant extraction is not only difficult because of the low yields but also because mixtures of highly similar compounds are often retrieved (Lehka *et al.* 2017). Alternative methods are required to produce sustainable and economical quantities of each compound. The flavonoid, resveratrol is believed to have many applications in the pharmaceutical and cosmetic industries (Liu *et al.* 2017). However, sustainable quantities of resveratrol cannot be chemically synthesised or extracted from any of its native organisms such as the Japanese knotweed (Mei *et al.* 2015). Sustainable resveratrol production was achieved by the development of an engineered *Saccharomyces cerevisiae* platform (Stenhuus *et al.* 2009). Similar microbial platforms have been achieved for some other phenolic compounds, vanillin being one example (Goldsmith *et al.* 2015).

PKs and NRPs are mainly found in bacteria and fungi; PKs are derived from intermediates in fatty acid biosynthesis to create complex cyclic structures whereas NRPs are produced by modular NRP synthetases. Most of these metabolites naturally function as antibiotics or antifungals. Both classes are of considerable interest to the pharmaceutical industry especially as antibiotics. Commercial PK antibiotic examples include the tetracyclines and macrolides while bacitracin and vancomycin are examples of NRPs. Other commercial applications of PKs and NRPs include immunosuppressive drugs, antivirals and anticancer treatments (Sun *et al.* 2006, Newman and Cragg 2012, Bloudoff and Schmeing 2017, Süssmuth and Mainz 2017).

The final major class of secondary metabolites are known as terpenes although there are several terpenes that are a part of primary metabolism. Terpenes are the most diverse group of metabolites with over 80000 structures currently characterised (Christianson 2017). Unlike the other secondary metabolites, all organisms contain one or both of the terpene biosynthesis pathways, namely the mevalonate and 2-methyl-D-erythritol-4-phosphate pathways (Lombard and Moreira 2010, Chandran *et al.* 2011, Wang *et al.* 2018). Terpenes have many industrial applications from being used as flavours and fragrances to antimalarial and anticancer treatments (Kirby and Keasling 2008, Arendt *et al.* 2016). However, similarly to alkaloids and phenolic compounds, terpenes are produced in low yields in the host organism resulting in an inefficient extraction process. Currently, most terpenes are extracted from plants which, as well as inefficient extraction, requires large areas of land for cultivation

(Caputi and Aprea 2011, Zebec *et al.* 2016). As many terpenes are in high demand by industries, many investigations have utilised microorganisms, predominantly *E. coli* and *S. cerevisiae*, to create terpene production platforms (Zebec *et al.* 2016, Paramasivan and Mutturi 2017). Several companies have succeeded in the industrial production of a handful of terpenes but further work is required to enable more terpenes to be made commercially available (Leavell *et al.* 2016).

1.2 Terpenes

The diversity of the terpene class of compounds is all derived from a five-carbon unit known as isoprene with a structure of C_5H_8 as described by Wallach (1887). In the mid-nineteenth century, Ruzicka (1953, 1959) described the "Biogenetic isoprene rule" that explained the mechanism by which all terpenes could be formed. This rule explains that isoprene units can be added to one another head-to-tail by electrophilic addition to form an extended, branched hydrocarbon chain of varying lengths. This prenyl chain could then be cyclised and further post-cyclisation modifications made to form the most diverse group of natural products. Since then, the activated forms of isoprene, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) have been identified as the five-carbon extender units. Initially, IPP and DMAPP are condensed head-to-tail to form the 10-carbon geranyl diphosphate (GPP). Further additions of IPP produce 15-carbon farnesyl diphosphate (FPP) and 20-carbon geranylgeranyl diphosphate (GGPP), with the loss of a diphosphate group as each unit is added. These additions are catalysed by the prenyltransferase enzymes, geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS) and geranylgeranyl diphosphate synthase (GGPPS) (Christianson 2007). Longer prenyl diphosphates can be produced by the condensation of further IPP units in cis or trans configurations or by head-to-head condensation (Liang *et al.* 2002, Tholl 2015). For example, the 30-carbon squalene is formed by head-to-head condensation of two FPP molecules (Lee and Poulter 2008). This study will focus on the shorter prenyl diphosphates and their cyclised products. For the shortest chains, the prenyl diphosphates are cyclised by terpene synthases (TPSs) to form hemiterpenes from IPP, monoterpenes from GPP, sesquiterpenes from FPP and diterpenes from GGPP (Figure 1.1). Traditionally, terpenes are the cyclised hydrocarbons while terpenoids are those structures with extra functional groups that can be added during the cyclisation by the TPS or post-cyclisation by tailoring enzymes including

hydroxylation, acetylation and glycosylation (Chang and Keasling 2006, Jiang *et al.* 2016, Wang *et al.* 2019). Here, all further cyclised products will be collectively referred to as terpenes.

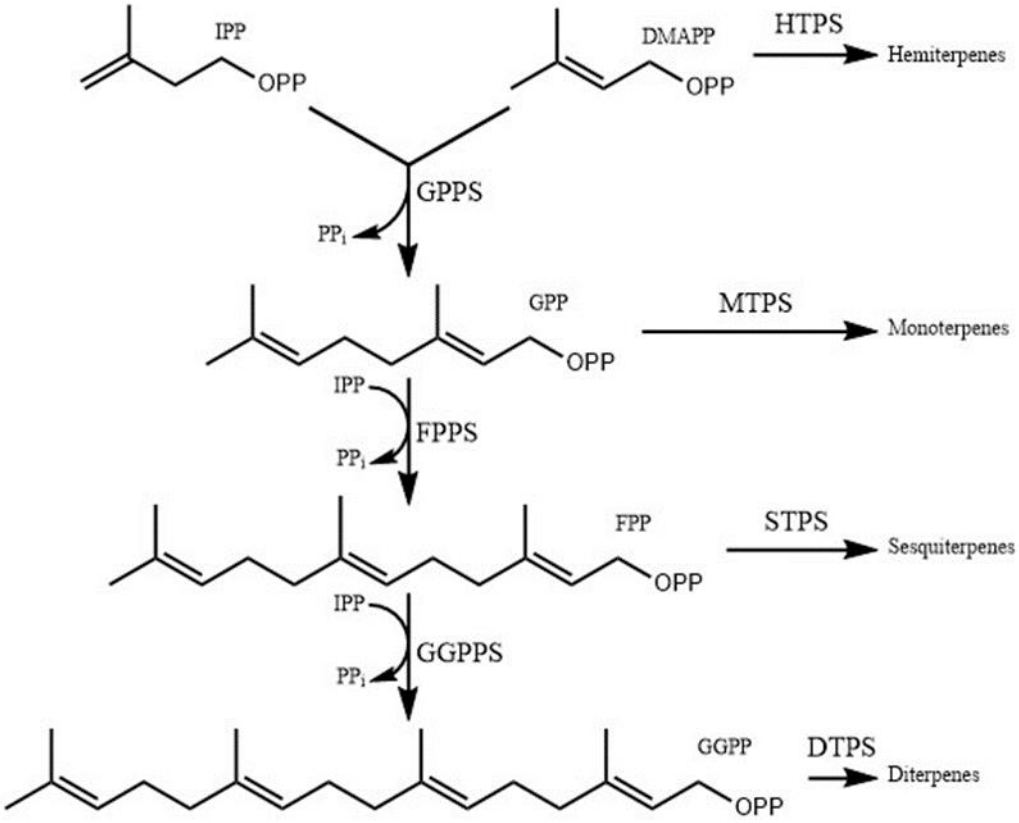


Figure 1.1: Biosynthesis of the short prenyl diphosphates by condensation of IPP and DMAPP followed by further IPP units being used to extend the chains. GPPS: Geranyl diphosphate synthase, GGPP: Geranylgeranyl diphosphate, HTPS: Hemiterpene synthase, P: Phosphate.

1.2.1 Terpene pathways

The two terpene biosynthesis pathways are the mevalonate (MVA) and 2-methyl-D-erythritol-4-phosphate (MEP) pathways which are responsible for the production of IPP and DMAPP in all organisms (Figure 1.2). The eight-stepped MEP pathway begins with the glycolysis intermediates, pyruvate and glyceraldehyde-3-phosphate and is used by the majority of bacteria (Boronat and Rodríguez-Concepción 2014). The classical MVA pathway contains seven steps from acetyl-CoA to IPP (Boronat and Rodríguez-Concepción 2014). This pathway is used by eukaryotes and most Archaea (Miziorko 2011, Nishimura *et al.* 2013). Several Archaea and bacteria, such as *Roseiflexus castenholzii* use variants of the MVA pathway that use different

enzymatic steps after mevalonate (Lombard and Moreira 2010, Dellas *et al.* 2013, Azami *et al.* 2014, Vannice *et al.* 2014, Vinokur *et al.* 2014). Plants are the only eukaryotes which contain both MVA and MEP pathways. The MEP pathway is localised to the plastids while the MVA pathway is found throughout the rest of the cell (Pulido *et al.* 2012). Some bacteria such as *Staphylococcus aureus*, do not contain the MEP pathway at all and instead only utilise the MVA pathway (Balibar *et al.* 2009, Pérez-Gill and Rodríguez-Concepción 2013). A variety of other bacterial species have also been shown to contain some or all of the enzymes involved in the mevalonate pathway (Lombard and Moreira 2010).

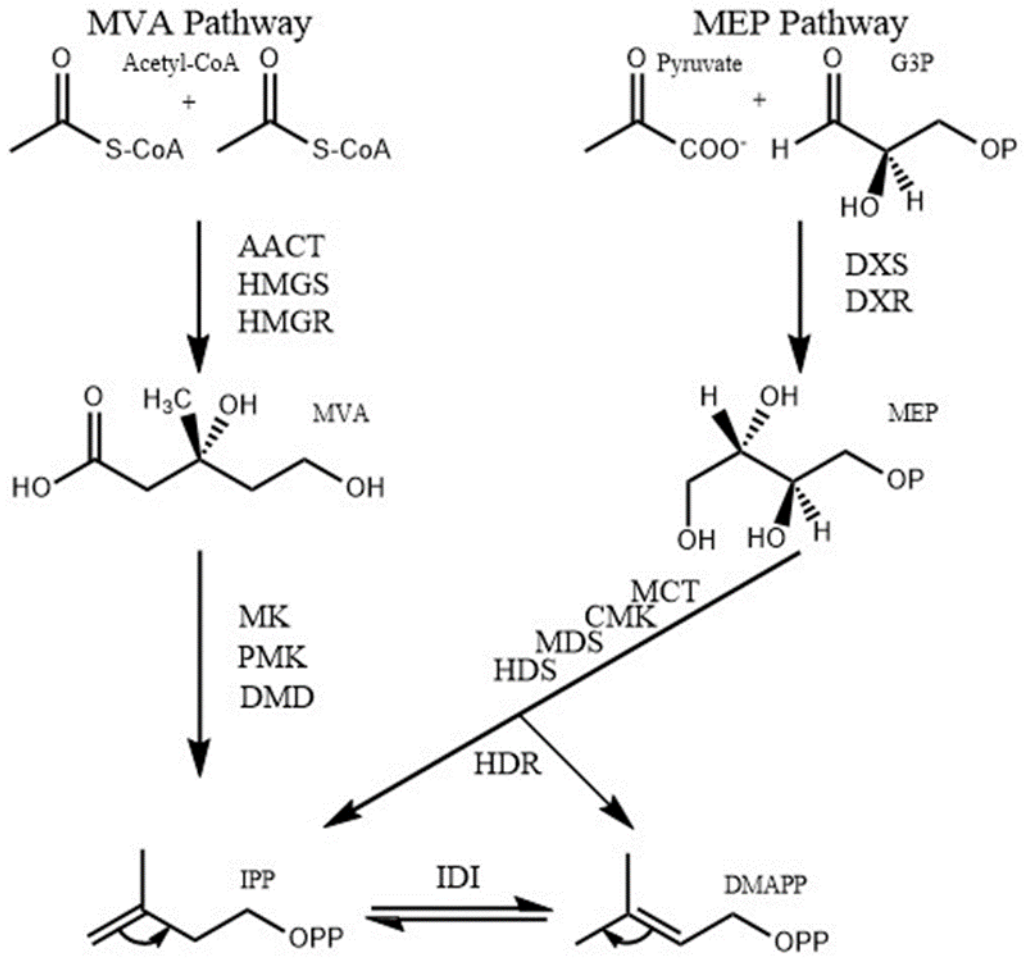


Figure 1.2: Classical mevalonate (MEV) and 2-methyl-D-erythritol-4-phosphate (MEP) pathways required for the biosynthesis of IPP and DMAPP for terpene production in all domains of life. See Boronat and Rodriguez-Concepción *(2014)* for more details about intermediate structures. Abbreviations: G3P: Glyceraldehyde-3-phosphate, DXS: 1-deoxy-D-xylulose-5-phosphate synthase, DXR: 1-deoxy-D-xylulose-5-phosphate reductase, MCT: 2-C-methyl-D-erythritol-4-phosphate Cytidylyltransferase, CMK: 4-diphosphocytidyl-2-C-methyl-D-erythritol Kinase, MDS: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate, HDS: (E)-4-hydroxy-3-methyl-but-2-enyl Diphosphate Synthase, HDR: (E)-4-hydroxy-3-methyl-but-2-enyl Diphosphate Reductase, P: phosphate.

1.2.2 Hemiterpenes

Hemiterpenes are the simplest terpenes and are synthesised from DMAPP and do not require any cyclisation. The only true hemiterpenes that use similar cyclase enzymes to the other terpene classes are isoprene and 2-methyl-3-buten-2-ol (Figure 1.3) but in these cases, these enzymes are only required to remove the diphosphate (Gray *et al.* 2011, Sharkey and Monson 2017). There are other hemiterpenes such as prenol and isoprenol but these are formed using alternative phosphatase enzymes (Zheng *et al.* 2013, Li *et al.* 2018). Isoprene and methylbutenol are formed by the removal of the diphosphate from DMAPP. The resulting carbocation either has a proton abstracted to form isoprene or a hydroxyl added to produce methylbutenol (Gray *et al.* 2011). Isoprene and methylbutenol are highly volatile compounds that are mainly produced by plants. Naturally, isoprene is predicted to be released by plant leaves to alleviate heat stress as well as preventing the accumulation of reactive oxygen species to avert any intracellular damage (Sharkey *et al.* 2008, Vickers *et al.* 2009). Methylbutenol is likely to perform similar functions (Gray *et al.* 2006). Isoprene released by plants has also been shown to repel herbivores and may act as a signalling molecule for growth (Laothawornkitkul *et al.* 2008, Jones *et al.* 2016). Isoprene has a wide range of applications; it is used in fragrances or as a chemical feedstock to form various polymers and may be used as a biofuel in the future (Bentley and Melis 2012, Ye *et al.* 2016). Isoprene is primarily used in the form of its polymer, polyisoprene, which is used to make synthetic rubber for numerous products (Ye *et al.* 2016).

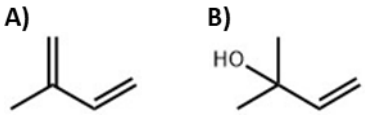


Figure 1.3: Structures of A) isoprene (2-methyl-1,3-butadiene), the empirical unit of all terpenes, and B) 2-methyl-3-buten-2-ol.

1.2.3 Monoterpenes

All of the terpenes produced by the cyclisation of GPP are known as the 10-carbon monoterpenes and may be linear, mono- or bicyclic structures (Figure 1.4). All cyclic monoterpenes go through an initial 1,6-cyclisation of the carbocation formed before further cyclisation and carbocation quenching (Zebec *et al.* 2016). Due to the

antibacterial and antifungal properties of most monoterpenes, the vast majority are produced by plants. Only two monoterpenes have been characterised as being produced by bacteria, 1,8-cineole and linalool (Nakano, Kim, *et al.* 2011a, 2011b) while only 1,8-cineole has been characterised from fungi (Shaw *et al.* 2015). The monoterpenes, sesquiterpenes and some diterpenes that are produced by plants are incorporated into volatile essential oils containing a variety of hydrophobic compounds, many of which are from the other secondary metabolite classes discussed previously. The diversity of terpene structures in these essential oils varies from species to species; many have been shown to have antimicrobial and insect repellent activity (Bakkali *et al.* 2008, Nerio *et al.* 2010, Kumari *et al.* 2014, Chouhan *et al.* 2017). The monoterpenes produced by plants have a variety of natural roles including allelopathy, as antimicrobials, repelling herbivorous organisms and attracting pollinators (Cheng and Cheng 2015, Herman *et al.* 2016, Krug *et al.* 2018, Lackus *et al.* 2018). For example, the production of the linear monoterpene alcohol, linalool, by *Arabidopsis thaliana* was shown to significantly reduce the number of aphids present on the plant (Aharoni *et al.* 2003). When the roots of *Populus nigra* and *Populus trichocarpa* were damaged by herbivores, a group of monoterpenes were released including 1,8-cineole and β-pinene. These monoterpenes have been shown to have inhibitory effects on growth of the plant pathogen, *Phytophtora cactorum* which causes root rot (Lackus *et al.* 2018). However, as with many terpenes, the ecological functions of many plant monoterpenes are still not known.

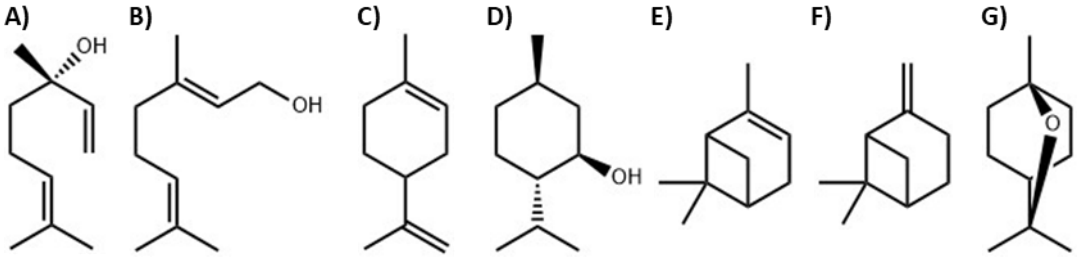


Figure 1.4: Examples of monoterpene structures. The linear monoterpenes are A) linalool and B) geraniol, monocyclic monoterpenes are C) limonene and D) menthol, and bicyclic monoterpenes are E) α-pinene, F) β-pinene and G) 1,8,cineole.

1.2.4 Sesquiterpenes

Sesquiterpenes are linear, mono-, bi- or tricyclic structures formed from the cyclisation of FPP (Figure 1.5). Whereas monoterpene structures are limited to an

initial carbocation 1,6-cyclisation, sesquiterpene biosynthesis can utilise 1,10- and 1,11-cyclisations with FPP and 1,6-, 1,7-, 1,10- and 1,11-cyclisations with nerolidyl diphosphate (NPP), which is formed when the diphosphate moiety reattaches itself to the initial carbocation at C3 rather than C1 (Dickschat 2016). As a result of more initial cyclisations being available, a more diverse range of structures can be produced. The majority of the characterised sesquiterpenes are isolated from plants although many have been characterised in bacteria and fungi. Like monoterpenes, many sesquiterpenes are also volatile, and therefore the known ecological functions overlap. As well as being used by plants to deter herbivorous insects (de Azevedo *et al.* 2003, Bleeker *et al.* 2010), sesquiterpenes can be used to attract predators of these insects or pollinators (Shields and Hildebrand 2001, Kappers *et al.* 2005, Schnee *et al.* 2006). For example, the sesquiterpene hydrocarbon, β-caryophyllene, is released by maize roots that have been wounded by larvae to attract nematode predators (Rasmann *et al.* 2005). It has also been shown to inhibit the growth of the bacterial plant pathogen, *Pseudomonas*, from growing on *A. thaliana* flowers (Huang *et al.* 2012). The biological functions of bacterial and fungal sesquiterpenes have not been investigated to the extent of plant sesquiterpenes (Kramer and Abraham 2012). Some fungal sesquiterpenes exhibit antibacterial and antifungal properties, and several are used to signal between species of fungi or between fungi and plants (Minerdi *et al.* 2011, Kramer and Abraham 2012). Pentalenolactone and albaflavenone are antibiotics produced by *Streptomyces* bacteria that are derived from the sesquiterpenes pentalenene and *epi*-isozizaene, respectively (Cane *et al.* 1990, Zhao *et al.* 2008).

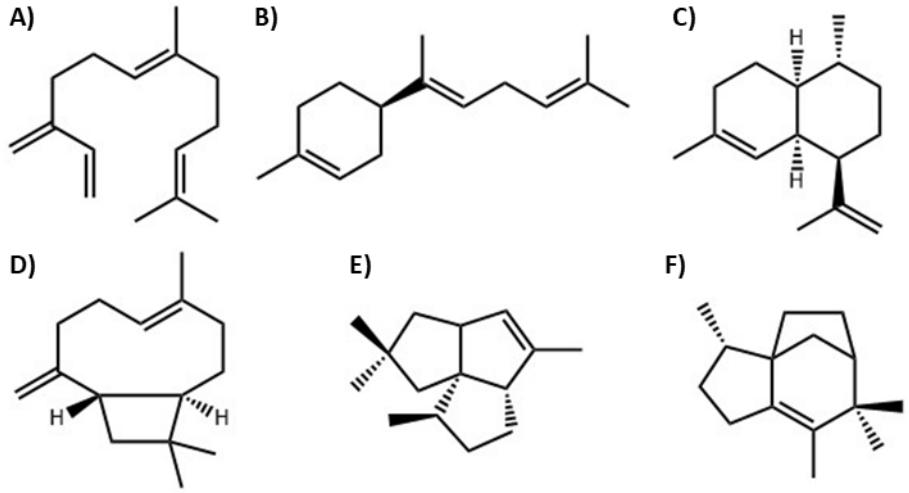


Figure 1.5: Examples of sesquiterpene structures. A) β-farnesene is a linear example while B) β-bisabolene is a monocyclic sesquiterpene. C) amorphadiene and D) β-caryophyllene are bicyclic sesquiterpenes and E) pentalenene and F) *epi*-isozizaene are tricyclic sesquiterpene structures.

1.2.5 Diterpenes

Diterpenes are 20-carbon terpenes formed from the cyclisation of GGPP (Figure 1.6). They are mainly isolated from plants but are also from bacteria and fungi. Unlike monoterpenes and sesquiterpenes, diterpenes are generally much less volatile due to their size, where they instead form resins in plants (Yáñez-Serrano *et al.* 2018, Liu *et al.* 2019). Diterpene structures can be linear, monocyclic, bicyclic, tricyclic, tetracyclic, pentacyclic or macrocyclic (Lanzotti 2013, Gong *et al.* 2014). The biological roles of diterpenes currently include as insect antifeedants, phytohormones, antimicrobials and protection from abiotic stresses such as oxidative stress (Munné-Bosch *et al.* 2001, Yamaguchi 2008, Zhao *et al.* 2011, Seo *et al.* 2012, Smanski *et al.* 2012, Pelot *et al.* 2018, Salazar-Cerezo *et al.* 2018). In plants, the production of gibberellin phytohormones is a vital part of primary metabolism for growth and development, however, bacterial and fungal plant pathogens can utilise gibberellins as virulence factors (Hamberger *et al.* 2011, Nagel *et al.* 2017). Many diterpenes produced by bacteria come from antibiotic gene clusters such as terpentecin and platencin (Hamano *et al.* 2002, Wang *et al.* 2007).

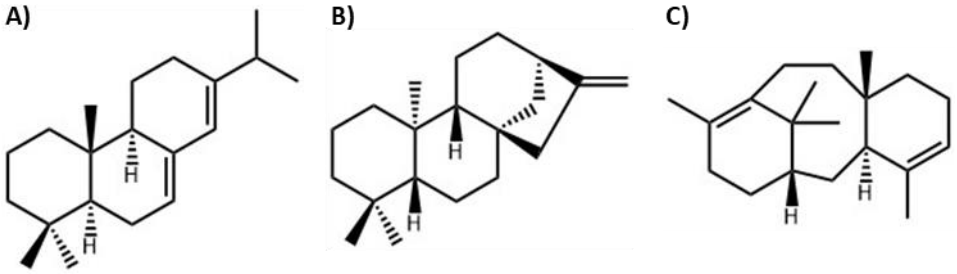


Figure 1.6: Examples of diterpene structures. A) Abietadiene is a precursor to dehydroabietic acid. B) ent-kaurene is the precursor of gibberellins in plants, bacteria and fungi. C) Taxadiene is the precursor to the anti-cancer drug, taxol produced by *Taxus* yew trees.

1.2.6 Commercial terpene production

Many terpenes have a variety of useful applications in industry and are in high demand. Currently, the primary method for producing most terpenes is to extract the terpene from plant material (Caputi and Aprea 2011, Weathers *et al.* 2011). Some terpenes can be isolated from waste feedstocks, for example limonene is extracted from orange peels, but most require large areas of land for cultivation of the producing plant (Durand and Durand 2009, Salehi *et al.* 2018, Jha *et al.* 2019). To maintain a continuous supply of terpene, this method relies on a consistent quantity of feedstock or harvest throughout the year which are known to fluctuate causing the price of the terpene to fluctuate also (Caputi and Aprea 2011). The terpenes are extracted from plants as essential oils which requires further purification to isolate each individual terpene. As the terpene yield is relatively low, this method is expensive and inefficient which is reflected in the price of the terpene product (Zebec *et al.* 2016). An alternative is to use microorganisms, such as *E. coli* and *S. cerevisiae*, to produce high yields of terpene as they are easier to engineer, cheaper and more efficient to enable demand to be met (Mehta *et al.* 2003, Schmidt *et al.* 2011, Gupta and Phulara 2015, Azhar *et al.* 2017, Liang *et al.* 2019). The two main methods for enabling heterologous terpene production in microorganisms is either to upregulate the native pathway or introduce a heterologous pathway.

Isoprene is currently the highest produced terpene, however almost all of it is petroleum-derived (Leavell *et al.* 2016). Multiple companies are investigating using microorganisms to provide sustainable platforms for terpene production including Amyris, Genencor and GlycosBio (Leavell *et al.* 2016). Genencor have successfully developed an isoprene platform using an engineered *E. coli* strain containing a

heterologous mevalonate pathway and the isoprene synthase from *Populus alba* to produce isoprene titres as high as 60 g/L (Whited *et al.* 2010, Chotani *et al.* 2013, Ye *et al.* 2016). The development of this *E. coli* system should enable sustainable isoprene production to replace petroleum as a source for the future.

As monoterpenes are volatile, many have unique aromas that are utilised in the fragrance industry such as geraniol and menthol which have rose- and mint-like aromas, respectively (Chen and Viljoen 2010, Caputi and Aprea 2011, Kamatou *et al.* 2013). As well as the flavour and fragrance industry, monoterpenes are used widely in the pharmaceutical and cosmetic industries (Caputi and Aprea 2011, Leavell *et al.* 2016). Limonene is a well-studied monoterpene that is utilised in cleaning products and the flavour and fragrance industry for its citrus taste and aroma. It has also been identified as having anticancer properties (Caputi and Aprea 2011, Jongedijk *et al.* 2016, Yang *et al.* 2017, Yu *et al.* 2018). Several monoterpenes are also being researched as biofuels or as additives to standards fuels including linalool, myrcene and α/β-pinene (Beller *et al.* 2015). The dimerization of α/β-pinene has been shown to have similar properties to that of JP-10 jet fuel (Meylemans *et al.* 2012, Beller *et al.* 2015).

Several studies have attempted to make monoterpenes in microorganisms. Engineered strains of *E. coli*, *S. cerevisiae* and *C. glutamicum* have been developed to produce a variety of monoterpenes (Zebec *et al.* 2016). This included the insertion of a heterologous mevalonate pathway into *E. coli*, over-expression of individual rate-limiting enzymes of the native pathway or manipulation of the native terpene pathway (Ignea *et al.* 2011, Du *et al.* 2014, Kang *et al.* 2014, Amiri *et al.* 2016). The highest productivity was ~200 mg/L/day of limonene using a proteomics based approach with an *E. coli* strain containing a heterologous mevalonate pathway with genes from *S. aureus*, *S. cerevisiae*, *Mentha spicata* and *Abies grandis* (Alonso-Gutierrez *et al.* 2013, 2015, Zebec *et al.* 2016). However, none of these systems produce high enough titres to be commercially available yet. Unlike most sesquiterpenes and diterpenes, this is believed to be caused by the toxicity of monoterpenes at higher concentrations due to their lipophilic properties that enable them to interact and disrupt the cell membrane (Brennan *et al.* 2012, Chacón *et al.* 2019, Demissie *et al.* 2019). Two methods have been developed to relieve the toxicity of monoterpenes by acetylating the hydroxyl of terpene alcohols or using cell-free systems (Korman *et al.* 2017, Chacón *et al.* 2019). Acetylating geraniol produced ~1.2 g/L/day of geranyl acetate, where the highest previous titre of geraniol was <0.1 g/L/day (Zhou *et al.* 2014, Chacón *et al.* 2019). The cell-free system was able to produce >1.4 g/L/day of

limonene, pinene and sabinene separately (Korman *et al.* 2017). The demand for menthol is around 30000-32000 metric tonnes per year while the demand of limonene was 45000 tons in 2015 which is expected to rise to 65000 tons by 2023 (Kamatou *et al.* 2013, John *et al.* 2017). As more applications are identified for monoterpenes and demand increases, development of these microbial methods to relieve toxicity should enable sustainable monoterpene production to enable the high industrial demand to be met.

Similarly, to monoterpenes, sesquiterpenes are being investigated for use or are currently used in the pharmaceutical, flavour and fragrance and biofuel industries, however, unlike monoterpenes, microbial production platforms have been developed for producing multiple industrially important sesquiterpenes. Artemisinin (Figure 1.7A) and several of its derivatives, originally identified from *Artemisia annua*, are used as antimalarials and have also been considered as anticancer therapies (Ridley 2002, Haynes 2006, Konstat-Korzenny *et al.* 2018). The demand of artemisinin cannot be met by the low yields produced when extracted from *A. annua* (Kumar *et al.* 2004, Hale *et al.* 2007, Weathers *et al.* 2011). To enable cheaper production of artemisinin, Amyris developed a semisynthetic microbial platform using an engineered *S. cerevisiae* containing an upregulated native mevalonate pathway with the amorphadiene synthase, amorphadiene oxidase and cytochrome P450 reductase genes from *A. annua*. The product, artemisinic acid, is extracted and then chemically converted to artemisinin (Westfall *et al.* 2012, Paddon *et al.* 2013). Similarly, the terpene compounds responsible for the fragrances of sandalwood and patchouli are commercially available from microbial platforms from Evolva and Isobionics as well as other companies (Albertsen *et al.* 2011, Leavell *et al.* 2016, van Beek and Joulain 2018). While Evolva use a similar system to Amyris in *S. cerevisiae*, Isobionics use an engineered *Rhodobacter sphaeroides* strain that utilises its own native MEP pathway as well as a heterologous mevalonate pathway (Huembelin *et al.* 2014, Orsi *et al.* 2019). Although sandalwood and patchouli oils are made up of multiple sesquiterpenes, only a single cyclase enzyme is required in each case to produce the woody smelling oils (Deguerry *et al.* 2006, Jones *et al.* 2011). The valuable flavour and fragrance compounds, valencene and nootkatone (Figure 1.7B and 1.7C) are also microbially produced by Evolva and Isobionics. Isobionics also produce β-elemene, β-bisabolene and δ-cadinene (Leavell, Mcphee and Paddon, 2016; http://www.isobionics.com/index.html). Valencene, nootkatone and β-elemene have an orange-like, grapefruit-like and ginger-like tastes and aromas, respectively (Beekwilder *et al.*, 2014; http://www.isobionics.com/index.html). Sesquiterpenes,

including β-farnesene, β-bisabolene, *epi*-isozizaene and α/β-cedrene have also been investigated for use as biofuels (Gupta and Phulara 2015, Harrison and Harvey 2017, Liu, Tian, *et al.* 2018). Of these sesquiterpenes, only β-farnesene is currently commercially produced by Amyris as a biofuel and chemical feedstock (Schofer *et al.* 2014, Meadows *et al.* 2016). The development of these microbial systems should enable other valuable sesquiterpenes to be produced in a renewable, efficient process by replacing the cyclase gene and any other downstream genes.

The biosynthesis of most of the monoterpenes and sesquiterpenes discussed above only require the cyclisation of the prenyl diphosphate produced by one of the pathways. Many diterpenes are more complex as they require up to two cyclase enzymes in addition to multiple post-cyclisation modifications to produce the final product. An example is the formation of the anticancer drug, taxol (Figure 1.7E), which is from the Taxus species of yew tree and is currently used to treat a wide range of cancers (Ajikumar *et al.* 2010, Khanna *et al.* 2015). Taxol was originally extracted from yew trees, however it required 2000-3000 fully grown trees, which can take up to 200 years, to isolate 1 kg of taxol which was not a sustainable production method (Kirby and Keasling 2008, Li *et al.* 2015). The biosynthesis of taxol from GGPP requires over 10 enzymatic steps including hydroxylations, acetylations, epoxidations as well as the initial cyclisation (Croteau *et al.* 2006, Li *et al.* 2015). This means that chemical synthesis of industrially useful or interesting diterpene compounds including taxol, dehydroabietic acid and steviol glycosides, is extremely challenging and not sustainable (Nicolaou *et al.* 1994, Trapp and Croteau 2001a, Brandle and Telmer 2007). The current method for producing taxol is to use plant cell culture of *Taxus chinensis* cells in large bioreactors as engineered *E. coli* and *S. cerevisiae* haven't been able to utilise the entire pathway yet (Bringi *et al.* 2003, Ajikumar *et al.* 2010, Ding *et al.* 2014, Li *et al.* 2015). Many diterpenes are being considered for use in the pharmaceutical industry as anti-cancer, anti-inflammatory, diabetes and AIDS treatments (Hanson 2009, Gong *et al.* 2014). Another commercially available diterpene is Ambrox® which has a sweet earthy aroma and is available by semisynthetic routes as an alternative to ambergris (Caniard *et al.* 2012). The intermediate sclareol (Figure 1.7D) is extracted from *Salvia sclarea* or produced by engineered *E. coli* before the remaining steps to Ambrox® are performed chemically (Schalk *et al.* 2012, Schalk 2013, Yang *et al.* 2016). Sclareol is also used in fragrances as well as for making other compounds in the fragrance industry (Bhatia *et al.* 2008, Gong *et al.* 2014). Whilst valuable diterpene compounds, like sclareol, that require minimal extra steps after cyclisation can be produced in commercially engineered

microbial systems, the ability to construct more complex pathways, such as those from plants, in microbes for terpene production is not yet possible.

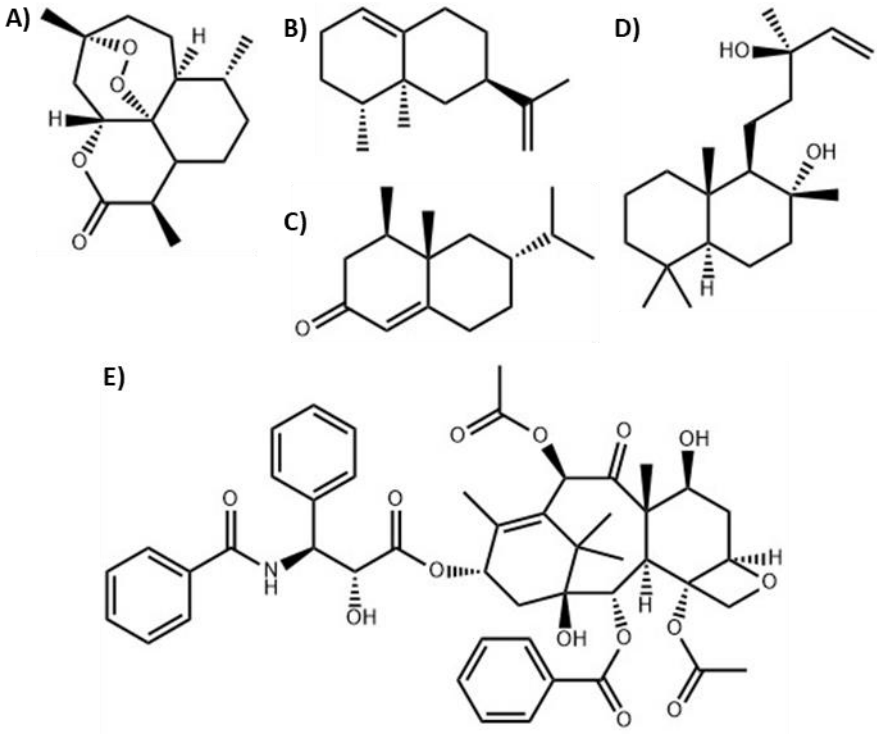


Figure 1.7: Structures of commercially available terpenes. A) The anti-malarial, artemisinin, the fragrance compounds, B) Valencene, C) Nootkatone and D) Sclareol and the anti-cancer drug E) Taxol.

1.3 Terpene synthase

The enzymes responsible for the diversity of terpene structures are the terpene synthases (TPSs). The function of these enzymes is to cyclise the prenyl diphosphates shown in Figure 1.1. Limonene cyclisation is shown in Figure 1.8A as an example. The TPSs discussed here only cyclise prenyl chains with *trans* configuration of double bonds, whereas there are other TPSs that cyclise prenyl chains with double bonds in the *cis* configuration (Matsuba *et al.* 2013, Beran *et al.* 2016, Christianson 2017). TPSs are globular enzymes that contain up to three α-helical domains (Diaz *et al.* 2011). These domains are called the γβα domains and always appear in this same order in the primary sequence of a TPS if present (Gao *et al.* 2013, Pazouki and Niinemetst 2016). The α-domain forms an α-helical bundle, also known as the α-fold, with the active site in the centre and TPSs that contain a functional α-domain are referred to as class I TPSs (Aaron *et al.* 2010, Baer, Rabe,

Fischer, *et al.* 2014). Class II TPSs contain a functional active site at the interface between the γβ domains that each form an α-barrel structure (Cao *et al.* 2010, Rudolf *et al.* 2016). The overall structural architecture generally varies depending on where the TPS originates from (Figure 1.8). All *trans*-forming prenyltransferases and microbial class I TPSs only contain the α domain (Liang *et al.* 2002, Gao *et al.* 2013). Plant class I TPSs all contain a functional α domain but, in addition, the β domain or the γβ domains may be present such as for limonene synthase and taxadiene synthase, respectively (Köksal *et al.* 2011, Morehouse *et al.* 2017). Where the latter domains are present, the class II active site between the γβ domains remains inactive (Köksal *et al.* 2011, Pemberton *et al.* 2017). Class II diterpene synthases from microbes and all triterpene synthases, such as squalene cyclase, have a γβ structure while plant class I and II diterpene synthases have a γβα tridomain structure (Wendt *et al.* 1997, Köksal *et al.* 2011, 2014, Rudolf *et al.* 2016). In most plant class II diterpene synthases, the α-domain is inactive (Köksal *et al.* 2014, Pelot *et al.* 2018). Protein structures of the varying domain architectures of TPSs are shown in Figure 1.9. There are multiple examples of bifunctional TPSs with either γβα or with rarer domain structures such as αα (Christianson 2017). For some TPSs that contain the γβα structure, the class I and class II active sites work together to produce a single terpene product. Examples include abietadiene synthase and aphidicolin-16-β-ol synthase where the class II active site catalyses the formation of copalyl diphosphate (CPP) and the class I active site then converts CPP to abietadiene and aphidicolin-16-β-ol, respectively (Peters *et al.* 2001, Hirte *et al.* 2018). There are several examples of functional TPSs that have an αα domain structure including the bacterial geosmin synthases and the fungal fusicoccadiene and ophiobolin F synthases (Toyomasu *et al.* 2007, Chiba *et al.* 2013, Harris *et al.* 2015). One of the α-domains of the latter fungal TPSs is a functional class I TPS while the other is a prenyltransferase. For the fusicoccadiene synthase, the prenyltransferase is a GGPP synthase while for the ophiobolin F synthase, it is a geranylfarnesyl diphosphate synthase. The class I TPS domains then cyclise the prenyl diphosphates subsequently synthesised by the prenyltransferase domains (Toyomasu *et al.* 2007, Chiba *et al.* 2013). Geosmin synthases are present in many *Streptomyces* species where both α domains are able to perform cyclisation reactions. The N-terminal domain cyclises FPP to germacradienol and the C-terminal cyclises and fragments germacradienol to geosmin and acetone (Jiang and Cane 2008, Harris *et al.* 2015).

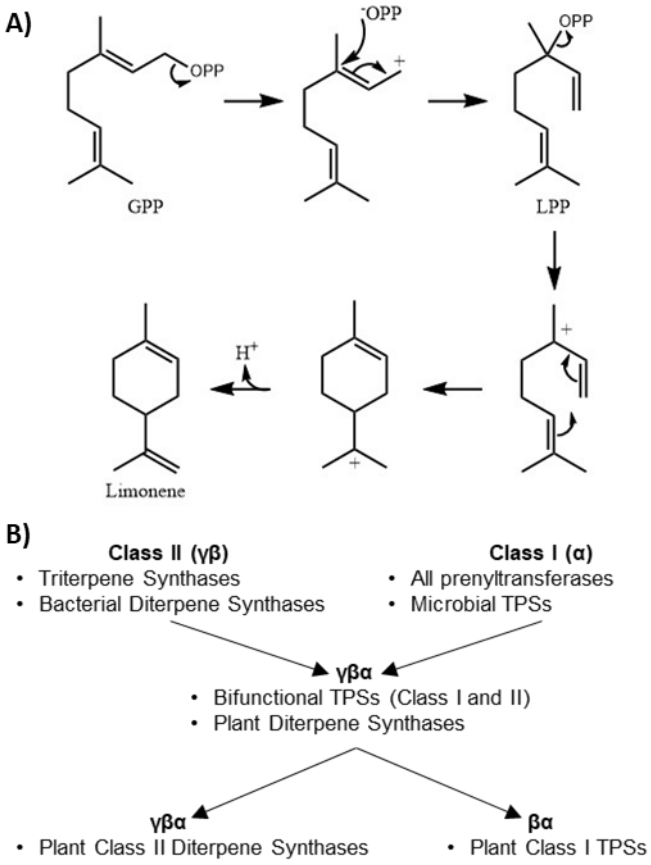


Figure 1.8: A) The predicted cyclisation of limonene from geranyl diphosphate (GPP) through the formation of the intermediate linalyl diphosphate (LPP) This image was adapted from Hyatt *et al*. (2007). B) Domain architecture of prenyltransferases, class I TPSs and class II TPSs. This image was adapted from Gao *et al. (2013)*.

To bind the prenyl chain, the active site cavities of both class I and class II are predominantly made up of hydrophobic residues. However, the TPS classes utilise different conserved motifs, cofactors and reaction mechanisms to catalyse cyclisation (Gao *et al.* 2013). The positions of these motifs are shown on the structures in Figure 1.9. Class I TPSs and prenyltransferases utilise three magnesium (Mg^{2+}) ions to coordinate the diphosphate moiety of the substrate in the active site (Liang *et al.* 2002, Hosfield *et al.* 2003, Köksal *et al.* 2010). The two conserved motifs that coordinate the Mg^{2+} ions in prenyltransferases are two aspartate-rich motifs with the sequence, DDXXD (Liang *et al.* 2002, Dickschat 2016, Christianson 2017). Class I TPSs are believed to have evolved from prenyltransferases where the second Asp-rich motif has become the NSE triad with a sequence of (N/D)D(L/I/V)XSX(K/R)(D/E) and the first Asp-rich motif has a conserved sequence of DDXX(X)D/E (Dickschat 2016, Christianson 2017). Most class I TPSs have an Asp-rich motif and an NSE triad, however there are variants in the sequence and even the presence of these motifs.

For example, the γ-humulene and δ-selinene synthases from *A. grandis* contain two Asp-rich motifs to catalyse cyclisation of FPP similar to the prenyl transferases while sodorifen synthase from *Serratia plymuthica* only has the first Asp-rich motif and no second conserved motif (Steele *et al.* 1998, Memari *et al.* 2013, Duell *et al.* 2019). Currently, the majority of bacterial class I TPS sequences have been characterised with the Asp-rich motif and the NSE triad (Yamada *et al.* 2015, Dickschat 2016).

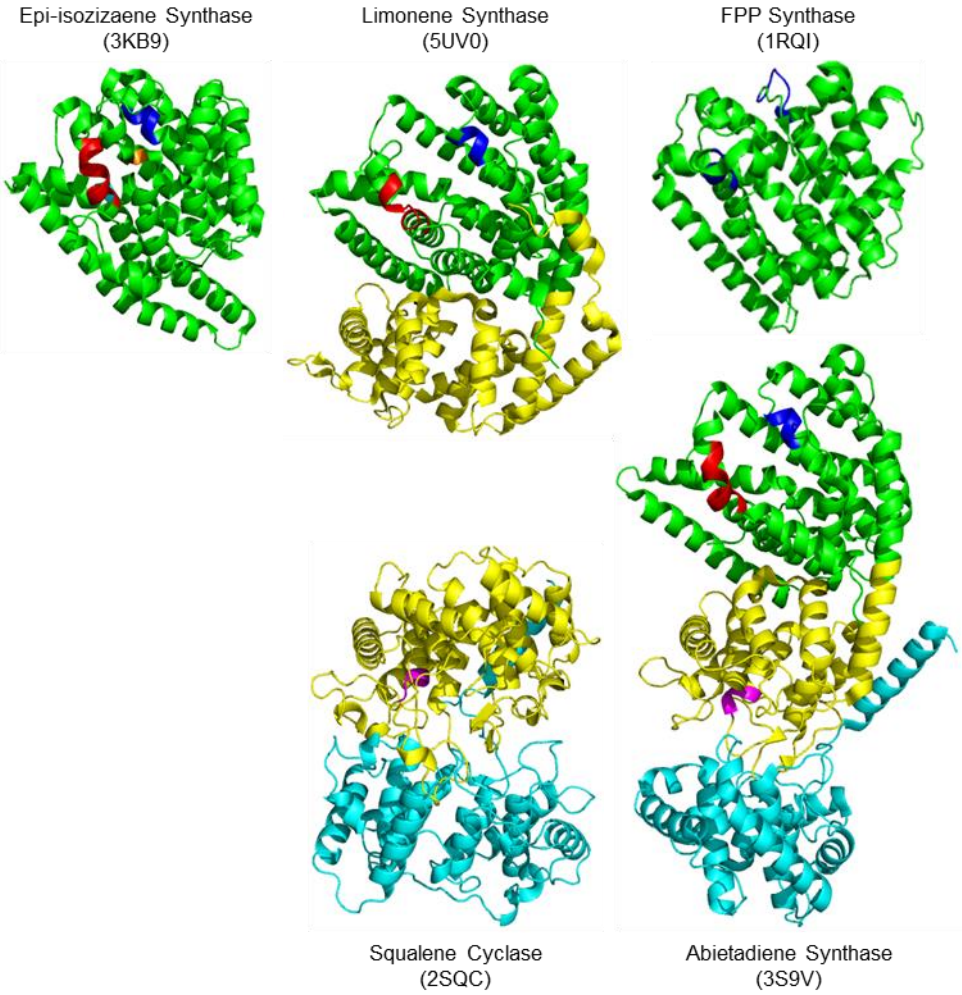


Figure 1.9: Variations of the domain architecture of the classes of TPSs and prenyl transferases. Class I TPSs contain the α-domain (Green) while class II TPSs contain the γβ domains (Cyan and yellow, respectively) as shown by *epi*-isozizaene synthase from Streptomyces coelicolor A3(2) and limonene synthase from Citrus sinensis. The classic motifs of a class I TPS are the Asp-rich (DDXXD) and NSE triad in red and blue, respectively. The extra motifs of a bacterial class I TPS are the pyrophosphate sensor (Cyan) and the RY motif (Orange). Although, the FPPS from E. coli has a class I α-helical structure, it has two Asp-rich motifs. Class II TPSs have an Asp-rich motif (DXDD) shown in magenta as shown by squalene synthase from Alicyclobacillus acidocaldarius and the abietadiene synthase from A. grandis. Both class I and II active sites are functional in abietadiene synthase. The PDB codes for each structure are in brackets and these structures are from Zhou et al., 2012, Aaron et al., 2010, Morehouse et al 2017, 2017, Zhang et al 2004 and Wendt, Lenhart and Schulz, 1999).

Prenyltransferases form the same α-helical structures as class I TPSs but the active site size varies depending on the length of the prenyl diphosphate being produced (Liang *et al.* 2002, Guo, Kuo, Chou, *et al.* 2004, Guo, Kuo, Ko, *et al.* 2004). The simplest way to discern one from the other is the presence of a second Asp-rich motif or an NSE triad (Liang *et al.* 2002, Dickschat 2016). Many conserved motifs required for function have been identified in class I plant TPSs, while only four conserved motifs, including the Asp-rich motif and the NSE triad, have been identified in class I bacterial TPSs (Fähnrich *et al.* 2014, Dickschat 2016). The other two are the pyrophosphate sensor and the RY motif identified by Baer et al (2014). None of the bacterial motifs or sequence identifiers enable mono-, sesqui- and diterpene synthases to be distinguished from one another (Yamada *et al.* 2012, 2015). Only in cases where there are large numbers of sequenced examples can they be identified such as geosmin, *epi*-isozizaene and 2-methylisoborneol synthases which occur widely in Streptomyces bacteria (Yamada *et al.* 2015). The RY motif is positioned at the top of the active site with the Asp-rich motif and NSE triad to aid coordination of the Mg^{2+} ions and diphosphate moiety of the substrate as well as closure of the active site to solvent (Aaron *et al.* 2010, Baer, Rabe, Fischer, *et al.* 2014, Rabe *et al.* 2016). On binding of the substrate, the prenyl chain is forced into a product-like conformation (Gao *et al.* 2013, Christianson 2017). By crystallising selinadiene synthase from *Streptomyces pristinaespiralis* with and without substrate analogues, Baer *et al.* (2014) were able to determine the function of the pyrophosphate sensor and the conformational changes to enable reaction initiation. The pyrophosphate sensor, which is an arginine residue, then moves towards the diphosphate of the substrate to hydrogen bond (Baer, Rabe, Fischer, *et al.* 2014, Rabe *et al.* 2016). This causes a further conformational change on the same helix of either an aspartate, threonine, serine or asparagine residue three positions away. This then puts the carbonyl oxygen of the next residue in close proximity to C3 of the prenyl diphosphate. This interaction allows the substrate to be ionised by enabling the diphosphate to leave resulting in the prenyl chain containing a highly reactive carbocation. The trio of residues to initiate cyclisation in selinadiene synthase from *S. pristinaespiralis* are R178, D181 and G182 (Baer, Rabe, Fischer, *et al.* 2014). Cyclisation of the prenyl chain involves the addition of the carbocation to carbon-carbon double bonds to enable cyclisation combined with hydride and alkyl shifts (Tantillo 2011, Rabe and Dickschat 2013, Dickschat 2016). The carbocation intermediates are stabilised by multiple interactions with the active site but mainly by π-cation interactions with aromatic amino acids (Aaron *et al.* 2010, Baer, Rabe, Fischer, *et al.* 2014). Most cyclisations terminate by the addition of a water molecule or by deprotonation. The

former creates a terpene alcohol while the latter forms a terpene hydrocarbon (Dickschat *et al.* 2014, Grundy *et al.* 2016). The only example of an alternative termination is for the monoterpene, bornyl diphosphate, from various plant species (Whittington *et al.* 2002, Despinasse *et al.* 2017). After cyclising GPP to the bicylic 2-bornyl cation, bornyl diphosphate synthase reattaches the diphosphate rather than deprotonating or adding a water to the carbocation (Weitman and Major 2010).

Class II TPSs contain only a single Asp-rich motif with the conserved sequence of DXDD in the β domain, which is independent of the class I Asp-rich motif, and utilises Mg^{2+} ions for the substrate binding or cyclisation (Cao *et al.* 2010, Smanski *et al.* 2012). However, unlike class I TPSs, class II TPSs can retain activity without the ion cofactors as they are not directly involved in the cyclisation mechanism and are used for binding the diphosphate (Prisic *et al.* 2007, Köksal *et al.* 2014). Similarly, to class I TPSs, the binding of the substrate places the prenyl chain in a product-like conformation (Köksal *et al.* 2014). However, rather than utilising the diphosphate of the substrate as a leaving group to form the reactive carbocation, the central aspartate of the Asp-rich motif protonates one of the carbon-carbon double bonds. A complex hydrogen bond network positions and stabilises the aspartate by the carbon-carbon double bond to enable it to donate a proton as a strong acid is required (Prisic *et al.* 2007, Köksal *et al.* 2014). The manipulation of the carbocation is similar to class I TPSs with various cyclisations, hydride shifts and methyl shifts followed by reaction termination by nucleophilic attack of a water molecule or proton abstraction (Peters 2010, Potter *et al.* 2014, 2016). The resulting cyclised prenyl diphosphate is further cyclised by a class I TPS which also removes the diphosphate (Hamano *et al.* 2002, Ikeda *et al.* 2007, Smanski *et al.* 2012). All characterised class II diterpene synthases have a class I diterpene synthase partner and sesquaterpene synthases that utilise 35-carbon prenyl chains also use class I TPS-class II TPS duos (Sato *et al.* 2011, Smanski *et al.* 2012, Gong *et al.* 2014). Similarly, triterpene synthases initiate cyclisation by protonation of a double bond in squalene or oxidosqualene but further cyclisation's are not performed by a class I TPS as there is no diphosphate moiety (Wang *et al.* 2010, Siedenburg and Jendrossek 2011, Tian *et al.* 2014).

The cyclisation performed by TPSs is one of the most complex biological reactions known and gives rise to the structural diversity of terpene structures. However, this complexity leads to promiscuity of the terpene product. The more carbocation intermediates involved in a mechanism, the more possible terpene structures that can be produced (Aaron *et al.* 2010). The active site of most TPSs is able to stabilise the many carbocation intermediates and direct them towards a terpene product with

specific regio- and stereochemistry preventing many unwanted stable intermediates from forming. For example, 91.5% of the terpene product for aristolochene synthase is aristolochene with germacrene A and valencene the two other products that are stabilised versions of one of the carbocation intermediates (Deligeorgopoulou and Allemann 2003). There are, however, some TPSs that show lower specificity and therefore higher promiscuity towards terpene formation. Several highly promiscuous TPSs are γ-humulene and δ-selinene synthases from *A. grandis* that produce 52 and 34 terpenes, respectively (Steele *et al.* 1998). Studies have been able to show that the specificity of a TPS can be manipulated to enable another terpene to be the major product or even to lower overall specificity (Deligeorgopoulou and Allemann 2003, Yoshikuni, Ferrin, *et al.* 2006, Aaron *et al.* 2010). For example, the specificity of γ-humulene synthase could be increased and switched to another terpene, such as β-farnesene, β-bisabolene or α-longipinene through minimal active site mutations and no adverse effects on the kinetics (Yoshikuni, Ferrin, *et al.* 2006). This promiscuity has been linked to the ability of TPSs to be able to produce different terpenes through minor changes in the active site contour (Yoshikuni, Ferrin, *et al.* 2006, Bian *et al.* 2017). As well as promiscuity with products, several TPSs are able to turnover more than one prenyl diphosphate substrate (Pazouki and Niinemetst 2016, Bian *et al.* 2017, Johnson *et al.* 2019). For example, one TPS from *Streptomyces clavuligerus* can utilise both GPP and FPP to linalool and nerolidol, respectively (Nakano, Kim, *et al.* 2011a).

1.4 The *Parageobacillus* terpene platform

Extremophiles are characterised as organisms that are able to grow in extreme conditions including high salt concentrations, high pressure, high or low pHs and high or low temperatures. Organisms that thrive at high temperatures are known as thermophiles. This can be split up further to thermophiles that have an optimal growth temperature of >50 °C while organisms that can grow above 80 °C are called hyperthermophiles (Vieille and Zeikus 2001, Turner *et al.* 2007, Zeldes *et al.* 2015). Thermophiles are mainly made up of archaeal and bacterial species with a small number of thermophilic fungi while hyperthermophiles are mostly archaea with some bacterial species (Maheshwari *et al.* 2000, Turner *et al.* 2007). These organisms grow in natural locations with high temperatures including hot springs, hydrothermal vents and deep-sea sediments (Vieille and Zeikus 2001, Reeve *et al.* 2016). Thermophilic

organisms have also been isolated from man-made industrial sites such as compost heaps, oil wells and waste sludge and wastewater plants (Vieille and Zeikus 2001, Fardeau *et al.* 2004, Sekiguchi *et al.* 2008, Tan and Ji 2010, Bosma *et al.* 2015).

Many commercial processes including the biofuel, pharmaceutical and food and drinks industries use mesophilic organisms or their enzymes (Zeldes *et al.* 2015). However, many industrial processes use extreme conditions that are not optimal for mesophilic organisms or their enzymes. For processes that require high temperatures, thermophiles and their enzymes provide attractive alternatives compared to their mesophilic variants. There are believed to be many advantages to using thermophiles or their enzymes including lower contamination risk from mesophiles, lower cooling costs, increased substrate solubility, higher reaction rates and they are more renewable and environmentally friendly than chemical processes (Vieille and Zeikus 2001, Zeldes *et al.* 2015). As well as being stable at high temperatures, thermophiles are also able to withstand low or high pHs, high salt concentrations and high pressure (Alain *et al.* 2002, Olsson *et al.* 2003, Mavromatis *et al.* 2009, Feyhl-Buska *et al.* 2016). Several industrially relevant thermophilic enzymes are commercially available through companies including DuPont, Novozymes and Verenium (Sarmiento *et al.* 2015). The most well-known commercial thermostable enzyme used in industry is the DNA polymerase from *Thermus aquaticus*, used in Polymerase Chain Reaction (PCR) for amplification of DNA sequences (Saiki *et al.* 1988). The remainder of the market for commercial thermostable enzymes is dominated by hydrolases such as proteinases, cellulases, amylases and lipases (Sarmiento *et al.* 2015, Rigoldi *et al.* 2018).

Individual enzymes are practical for industrial processes that only require one or two steps. For more complicated processes, using a microorganism that can encode entire pathways would result in a more efficient process. Even though many thermophiles have shown promise for becoming platforms for natural product production, none are currently used commercially. For an organism to be competitive with these model platforms, it needs to be cost effective. To enable optimal titres of the desired product, the tools to genetically engineer a thermophile are required. This includes being able to use either transformation or conjugation, easily manipulate the genome, having a range of characterised promoters of varying strengths, having a range of selection markers and multiple origins of replication for multiple plasmids to be used for shuttling in DNA. To increase titres of the desired product either the native pathways can be upregulated, heterologous gene sequences can be inserted or

competing pathways can be deleted to enable a platform to be commercially viable and competitive.

As discussed previously, the commercial terpene platforms used by Amyris, Evolva and Isobionics use engineered *S. cerevisiae* and *R. sphaeroides* (Hansen 2011, Paddon *et al.* 2013, Huembelin *et al.* 2014). Recently, the first engineered terpene platform was described in the thermophile, *Parageobacillus thermoglucosidasius*. The *Parageobacilli* and *Geobacilli* species grow optimally between 40-70 °C and are gram-positive, aerobes or facultative anaerobes (Cripps *et al.* 2009, Zeigler 2014). *P. thermoglucosidasius* in particular has the potential to be an interesting target for natural product biosynthesis as it is able to utilize both pentose and hexose sugars, take up oligomeric sugars as well as having the appropriate genetic tools to enable comprehensive genetic engineering (Cripps *et al.* 2009, Reeve *et al.* 2016, Mougiakos *et al.* 2017). These sugars are produced from the breakdown of lignocellulosic biomass, where the *Parageobacilli* genus uses mixed acid fermentation to convert the sugar to ethanol, lactate, formate, succinate and acetate (Tang *et al.* 2009, Hussein *et al.* 2015). As *Parageobacilli* are able to produce many products that are useful industrially, they can be engineered to overproduce them. For example, a *P. thermoglucosidasius* platform was engineered for ethanol production by upregulating pyruvate dehydrogenase and deleting competing fermentation pathways for the formation of lactate and formate. This strain was known as *Parageobacillus thermoglucosidasius* TM242 (Cripps *et al.* 2009). *P. thermoglucosidasius* has also been engineered to overproduce isobutanol with yields up to 3.3 g/L. This system used an engineered isobutanol pathway on a plasmid containing native genes and genes from other species of the *Bacilli* class under the lactate dehydrogenase promoter from *Geobacillus thermodenitrificans* (Lin *et al.* 2014).

To produce terpenes in *Parageobacillus*, either the native MEP pathway would require upregulating or a heterologous mevalonate pathway introduced. As discussed previously, studies looking at the manipulation of the native MEP pathway have shown that terpene titres have never been as high as using an engineered heterologous mevalonate pathway. To upregulate and balance the native MEP pathway would require a range of promoters to prevent the build-up of any intermediates. Alternatively, a heterologous mevalonate pathway produces non-native intermediates that cannot be incorporated into microbial processes except for terpene production and cannot be regulated by the host organism. The main problem with building a heterologous mevalonate pathway in *P. thermoglucosidasius* is that the mevalonate pathway is mostly native to eukaryotes and archaea. Very few

bacteria contain the mevalonate pathway and none of them are thermophilic (Wilding *et al.* 2000, Balibar *et al.* 2009). Styles *et al.* (To be published) instead sourced most of the mevalonate pathway from the archaeon, *Sulfolobus solfataricus* P2. These genes were the *aact*, *hmgs*, *hmgr*, *mk*, *pmk*, *dmd* and *idi* (Kim *et al.* 1999, Yamashita *et al.* 2004, Nishimura *et al.* 2013). AACT and HMGS had not been characterized previously in *S. solfataricus* P2. The *hmgs* and *aact* gene sequences from the *S. solfataricus* P2 genome were recognized by homology searches. As the two steps producing IPP and FPP are present in both pathways, thermophilic bacterial variants of the *idi* and *fpps* genes could have been used. The *fpps* used came from *Geobacillus stearothermophilus* (Koyama *et al.* 1993). To determine whether the heterologous mevalonate pathway was functional, a thermostable TPS was required to turn over the FPP into a measurable terpene. The only characterised naturally thermostable TPSs were the τ-muurolol synthases from RoseRS_3509 and Rcas_0662 from *Roseiflexus* sp. RS-1 and *Roseiflexus castenholzii* DSM 13941 (Styles *et al.* 2017). RoseRS_3509 was used as a reporter for the pathway. The whole pathway was assembled into three operons in the pG2K_oriT_Mev9 plasmid (Figure 1.10). All of the operons were under the control of individual maltose-inducible promoters, aka the *pglv* promoter (Chia and Leak, unpublished). The first operon only contained *roseRS_3509*, with the second containing *hmgr*, *aact* and *hmgs* and the third containing *mk*, *dmd*, *pmk*, *idi* and *fpps*.

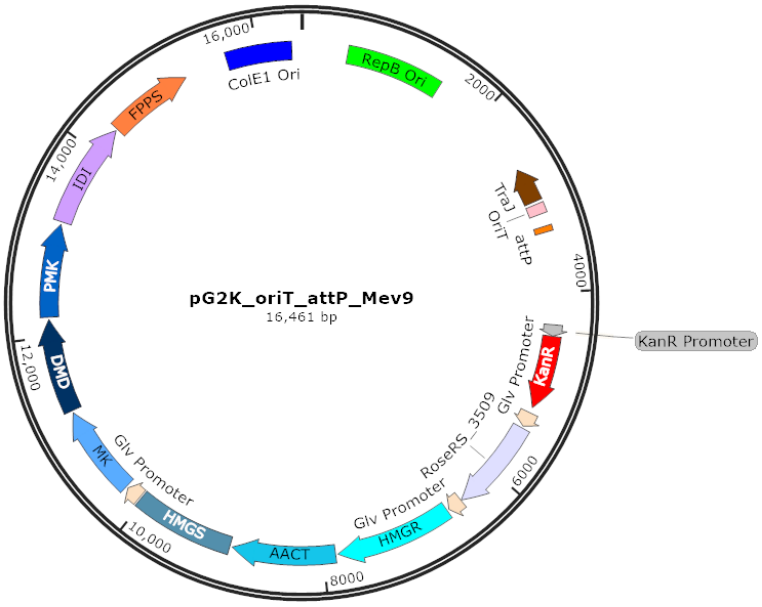


Figure 1.10: Plasmid map of pG2K oriT attP Mev9. This vector was used to enable the heterologous mevalonate pathway to be expressed in *P. thermoglucosidasius* NCIMB 11955. The plasmid pG2K oriT Mev9 was identical other than it lacked the attP site. The oriT was the origin of transfer to enable this plasmid to be conjugated from *E. coli* into *Parageobacillus* and the attP allowed for integration into the *Parageobacillus* chromosome if it contained an attB site and a serine integrase was present. This vector was constructed by Dr Matthew Styles (University of Bath, UK).

The pG2K_Mev9 plasmid was initially transformed into *E. coli* C43(DE3), where 5.0 mg/L of τ-muurolol was produced under induction with 1% maltose over 24 hrs. This construct produced more τ-muurolol than a control containing only HMGR, ACAT, HMGS and RoseRS_3509 which only produced 0.3 mg/L τ-muurolol. This provided evidence that the pathway was functional and was therefore conjugated into *P. thermoglucosidasius* NCIMB 11955. Over 48 hours, this strain produced minimal τ-muurolol and grew slowly suggesting that the pathway was causing some toxicity. As all of the genes were expressed using the same promoter, this likely leads to unbalanced pathway as each enzyme has a different rate of activity meaning that toxic intermediates, such as HMG-CoA, could build up (Pitera *et al.* 2007, Kizer *et al.* 2008). To relieve toxicity, an attP site was introduced to enable integration into the *P. thermoglucosidasius* NCIMB 11955 genome using a serine integrase. Although this chromosomal Mev9 (cMev9) strain still grew slowly, inducing with 1% maltose resulted in five times more τ-muurolol to be produced suggesting that lowering the expression levels of the genes enabled higher τ-muurolol titres to be produced.

To test this system, waste bread was used as a feedstock as *Parageobacillus thermoglucosidasius* secretes an α-amylase that can break down the starch from the

bread to maltose. The maltose released in turn auto-induces the expression of the pathway. To test this, the *P. thermoglucosidasius* NCIMB 11955 cMev9 strain was grown on 1% w/v bread. This produced similar τ-muurolol titres to when 1% maltose was used for induction. However, when 2% w/v bread was used, this resulted in eight times more τ-muurolol being produced than when using 1% w/v bread. To develop this system further, the remaining enzymes encoded should be fully characterised to enable the pathway to be balanced. This might require identifying and characterising other thermostable homologues to enable the optimal pathway to be built. Also, inducible promoters of varying strengths would need to be made using a similar method to Reeve *et al.* (2016) with the constitutive *prpls* promoters. Once this has been achieved, the conditions to optimise terpene production in *Parageobacillus* can be fully investigated. As sesquiterpenes are likely to be volatile at temperatures above 50 ^{o}C, a closed, continuous system could be developed where the terpene is distilled directly from the culture. The development of this thermostable pathway shows that sesquiterpenes can be produced at high temperatures. As well as improving the pathway towards higher τ-muurolol titres, more thermostable TPSs need to be identified to increase the variety of terpenes that can be produced at high temperatures.

1.5 Aims and Objectives

The primary aim of this project was to make a thermophilic terpene production platform more viable by increasing the number of terpenes that could be produced at high temperatures. As few naturally thermostable TPS have been isolated, the initial goal was to use developed search methods to find and characterise more of these enzymes from thermophile genome sequences. The second goal was to use protein engineering to manipulate the properties of this enzyme group to be more robust at high temperatures. Previous investigations into increasing the thermostability of this enzyme group mainly focussed on using directed evolution due to a lack of availability of thermostable models or structures. This study aimed to use the thermostable TPSs, RoseRS_3509 and Rcas_0662, as models to investigate thermostability of this enzyme group in an effort to try and determine stabilising mutations. This was combined with using computational methods to increase thermostability of various TPSs. Thirdly, mutations were inserted into the active site of the thermostable τ-muurolol synthase, RoseRS_3509, to alter the product distribution towards forming other terpenes as the dominant product. The initial target of these mutations was towards β-farnesene. Put simply, these strategies were to:

- Search and characterise new thermostable TPSs from thermophiles using Hidden Markov Models.
- Identify mutations that increase the thermostability of mesostable tau-muurolol TPSs using rational design with the thermostable tau-muurolol synthase and directed evolution while maintaining terpene production at higher temperatures.
- Target the production of farnesene by altering the active site of the thermostable tau-muurolol synthase, RoseRS_3509.
- Express any new thermostable TPSs in *Parageobacillus thermoglucosidasius* NCIMB 11955 with a heterologous mevalonate pathway to test terpene titres.

2 Materials and Methods

All reagents were sourced from Sigma-Aldrich (UK), Thermo Scientific (UK) and Merck Millipore (UK) unless stated otherwise. All solutions were made up in deionised water or Milli-Q® water. Solutions were sterilised, when necessary, by filter-sterilisation or autoclave (121 ^{o}C and 15 psi for 30 mins).

2.1 DNA Manipulation

2.1.1 Plasmid DNA

Antibiotic concentrations used for *E. coli* were 100 µg/ml and 50 µg/ml for ampicillin (Amp) and kanamycin (Kan), respectively. For *Parageobacillus thermoglucosidasius* NCIMB 11955, chloramphenicol (Chl) and Kan were used at 5 µg/ml and 12.5 µg/ml, respectively. The full list of plasmids used and created in this work is shown in Appendix Table 2. The initial plasmids used are shown in Table 2.1. The plasmid maps for P2 UM3 and P2 LM5 are shown in Appendix Figure 10 while the plasmid backbones of pG2K oriT SDM bgl sf-gfp GGready and pG2AC oriT SDM are shown in Appendix Figure 9.

Table 2.1: All of the initial plasmids used in this work. The description and with the associated antibiotic resistance genes and whether the plasmids can be sued for *E. coli* (Ec) or *Parageobacillus thermoglucosidasius* NCIMB 11955 (Pt) are included. MVO: Kindly provided by Maria Vittoria Ortenzi. MQS: Kindly provided by Dr Matthew Styles.

Plasmid	Description	Antibiotic	Source
pET28a	Expression vector with His-tag	Kan^R (Ec)	Novagen (USA)
pMA	Holding Vector for genes ordered from GeneArt	Amp^R (Ec)	Invitrogen (UK)
P2 UM3	Vector containing the upper mevalonate pathway genes; AACT, HMGS, and HMGR	Kan^R (Ec/Pt)	MQS

P2 LM5	Vector containing the lower mevalonate pathway; MK, PMK, DMD, IDI and FPPS	Kan^R (Ec/Pt)	MQS
pG2AC oriT SDM	Conjugation vector between *E. coli* and *Parageobacillus*	Amp^R (Ec), Chl^R (Pt)	MVO
pG2K oriT SDM bgl sf-gfp GGready	Golden Gate vector that enabled promoter-gene constructs to be built to add to pG2AC plasmid.	Kan^R (Ec/Pt)	MVO
pCR™-Blunt	Blunt ends to enable cloning of PCR or blunt fragments	Kan^R (Ec), Zeo^R	Thermo Fisher Scientific (UK)

2.1.2 Polymerase Chain Reaction (PCR)

Reactions were made up of 4.0 µl 5x Phusion HF buffer, 1.0 µl dNTPs (10 mM each), 1.0 µl 10 µM Forward primer, 1.0 µl 10 µM Reverse primer, 0.2 µl plasmid DNA, 1.0 µl DMSO, 0.2 µl 2 U/µl Phusion Hot Start II DNA Polymerase and 11.8 µl Milli-Q® water. For negative controls, plasmid DNA was replaced with Milli-Q® water. Each PCR was performed in 0.25 ml PCR tubes and controlled by a Mastercycler Gradient Thermal Cycler (Eppendorf, UK) and the program for all PCR reactions is shown in Table 4.

Table 2.2: Thermocycler program run for all PCR reactions in this study. The annealing temperature was dependent on the melting temperature of the primer sets while the extension time depended on the length of the fragment being amplified. [1] The annealing temperature was 5 °C below the lowest T_m of the primer pair as calculated by OligoAnalyzer 3.1. [2] Extension times were decided by the length of the amplicon using 30 secs/kb.

Stage	Cycles	Temperature (°C)	Duration
Denaturation	1	98	2-5 mins
Amplification	30-35	98	30 secs
		[1]	30 secs
		72	[2]
Extension	1	72	5-10 mins
Cooling	--	4	∞

2.1.3 Golden Gate Cloning

Golden Gate cloning was adapted from Engler (2008, 2009). PCR primers were designed to amplify inserts with 4 bp overhangs a single base from the BsaI restriction site. These overhangs enabled the inserts to be placed in the correct order and orientation in the vector. All BsaI restriction sites were positioned so that they were removed from the final vector after Golden Gate cloning. Reactions contained 100 ng plasmid backbone, equimolar concentrations to the plasmid backbone of each insert, 1.5 µl T4 DNA Ligase Buffer, 1.0 µl FastDigest Eco31I (FastDigest BsaI), 2.0 µl T4 DNA Ligase (1 million units/ml) and Milli-Q® water up to 15.0 µl. The program for the thermocycler is shown in Table 5.

Table 2.3: Thermocycler program used for Golden Gate cloning.

Stage	Cycles	Temperature (ºC)	Duration (mins)
Initial Restriction Digestion	1	37	30
Restriction Digestion	50	37	3
Ligation		16	4
Elimination Restriction Digestion	1	50	5
Inactivation	1	80	5
Cooling	--	4	∞

2.1.4 Colony PCR (cPCR)

Chosen colonies from transformation plates were spotted on LB agar replica plates for *E. coli* or 2YT agar replica plates for *P. thermoglucosidasius* NCIMB 11955 before resuspending them in 7 µl Milli-Q® water. *P. thermoglucosidasius* colony resuspensions were boiled at 99 ºC for 10 mins, frozen at -80 ºC for 10 mins and then thaw at room temperature to aid cell lysis. Each 20 µl reaction contained the resuspended colony, 1.0 µl 10 µM Forward primer, 1.0 µl 10 µM Reverse primer, 1.0 µl DMSO and 10.0 µl 2x MyTaq™ Red Mix (Bioline, UK). All cPCR reactions were thermally controlled by the thermocycler with the cPCR program shown in Table 6.

Table 2.4: Thermocycler program used to identify positive colonies by cPCR. [1] The annealing temperature was determined by the lowest melting temperature of a primer pair and the extension time by the length of the amplified fragment. [2] Extension times were decided by the length of the amplicon using 30 secs/kb.

Stage	Cycles	Temperature (°C)	Duration
Denaturation	1	95	5 mins
Amplification	30-35	95	30 secs
		[1]	30-45 secs
		72	[2]
Extension	1	72	5-10 mins
Cooling	--	4	∞

2.1.5 Primers

The online tool, OligoAnalyzer 3.1 (Integrated DNA Technologies, USA), was used to design all primers and these were synthesised by Eurofins Genomics (Germany). The primers used in this study are shown in Appendix Table 1.

2.1.6 DNA Agarose Gel Electrophoresis

DNA fragments from (c)PCR and restriction digestion reactions were separated using 1% (w/v) agarose gels. Agarose was dissolved in 1x TAE buffer, diluted from 50x TAE buffer containing 2 M Tris, 1 M glacial acetic acid and 64 mM EDTA. After heating, SYBR® Safe DNA stain was added to a concentration of 1x from 10000x. The marker, GeneRuler™ 1kb DNA Ladder (Figure 2.1), was added (7 µl) to one lane to determine DNA fragment size. All PCR samples had 6x DNA Gel Loading Dye added prior to being load onto the gel. Gels were run in Wide Mini-Sub® Cell GT Cell (Bio-Rad, UK) with a PowerPac™ Basic (Bio-Rad, UK) containing 1x TAE buffer at 120V for 40 mins. All gel images were taken using the G:Box (Syngene, UK) using the Gene Snap software (Version 7.09.17, Syngene, UK).

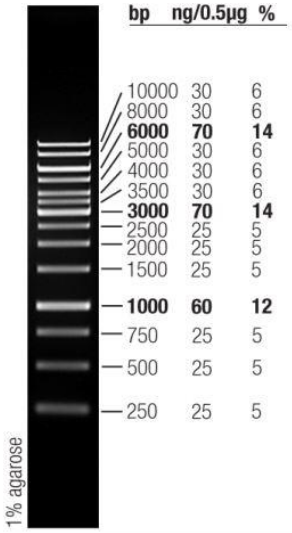


Figure 2.1: GeneRuler™ 1kb DNA Ladder used to determine DNA fragment size on agarose gels.

2.1.7 DNA Purification

2.1.7.1 Gel Extraction

Gel slices were visualized using a UV box (Fotodyne, USA) or a blue light illuminator (Cleaver Bioscience Ltd, UK) and DNA fragments excised with a scalpel. Following the manufacturers protocol, the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA) was used to extract PCR fragments or restriction digested DNA from agarose gel slices. Elution of DNA from the column used 10 µl of Milli-Q® water with a 1 min incubation at room temperature before centrifuging. All DNA was stored at -20^{o}C.

2.1.7.2 Plasmid Isolation

Following the manufacturers protocol, the QIAprep® Spin MiniPrep Kit (Qiagen, UK) was used to purify plasmid DNA from overnight *E. coli* cultures in 10 ml LB media. Plasmid DNA was eluted with 50 µl Milli-Q® water and stored at -20 ^{o}C.

2.1.8 Restriction Endonuclease Digestion

Restriction digestions were performed in 20 µl or 50 µl volumes in 0.25 ml PCR tubes. Manufacturers protocol was followed as reactions contained FastDigest enzymes (Thermo Fisher Scientific, UK). Plasmid DNA was digested at 50.0 ng/µl and 6.7 ng/µl for purified PCR fragments. Digested fragments were separated by DNA Agarose Gel Electrophoresis and isolated by gel extraction.

2.1.9 Gibson Assembly

Gibson Assembly protocol was adapted from (Gibson *et al.* 2009). DNA fragments were PCR amplified with 40 bp complementary overhangs at both 5' and 3' ends. In each reaction, 100 ng of the largest fragment was added with equimolar quantities of each fragment in a total volume of 5 µl. The master mix was made up of 320 µl 5X ISO Buffer (0.5 M Tris-HCl pH 7.5, 50 mM $MgCl_2$, 1 mM dNTP mix, 50 mM DTT, 25 % PEG-8000, 5 mM NAD), 0.64 µl 10 U/µl T5 exonuclease (NEB), 2 U/µl Phusion polymerase (NEB), 40 U/µl Taq DNA Ligase and Milli-Q® water up to 1.2 ml. The components of each reaction were added to 15 µl aliquots of master mix before heating to 50 ^{o}C for 60 mins in a thermocycler. For each Gibson Assembly reaction, 5 µl was added to a chemically competent *E. coli* aliquot for transformation.

2.1.10 Ligation

All ligation reactions were transformed into chemically competent *E. coli* Bio*Blue*.

2.1.10.1 Blunt End Cloning

The manufacturers guide of the Zero Blunt® PCR Cloning Kit (Invitrogen) was used to ligate blunt genes into pCR®-Blunt vector. This was used for the JKG1, FJSC11 and Tchrom genes that were synthesised as blunt.

2.1.10.2 Sticky End Cloning

All ligation reactions were incubated at room temperature for between 15 mins to overnight and had an insert:vector ratio of 3:1. In accordance with the manufacturers protocol, each reaction consisted of 0.5-3.0 µl restricted PCR amplified DNA fragment, 1.0 µl restricted pET28a, 1.0 µl 10x T4 DNA Ligase buffer, 1.0 µl T4 DNA Ligase (5 U/µl) and made up to 10.0 µl with Milli-Q® water.

2.1.11 Site Directed Mutagenesis

2.1.11.1 Overlapping PCR

To introduce a mutation into a gene, two primer pairs were used. The first pair contained a forward primer at the 5' of the gene paired with an internal reverse primer. The second pair contained an internal forward primer with a reverse primer 3' of the gene. The internal primers had ~10 bp overlap which contained the desired mutation in the centre of this overlap. The PCR was performed as in chapter 2.4.2 with both primer pairs to produce two fragments isolated using DNA agarose gel electrophoresis and gel extraction. A second PCR reaction was performed with the overlapping fragments. The program is shown in Table 2.5 and the reaction contained 4.0 µl 5x Phusion HF buffer, 0.4 µl dNTPs (10 mM each), 0.5 µl PCR amplified upstream DNA fragment (100 ng), 0.5 µl PCR amplified downstream DNA fragment (100 ng), 0.2 µl Phusion Hot Start II DNA Polymerase (2 U/µl) and Milli-Q® water up to 15 µl. This reaction allows the overlapping fragments to anneal to create a full-length mutant gene. After the first set of cycles, the program is paused to allow the addition of a 5 µl master mix containing 1 µl of each of the forward and reverse primers (10 µM), 1 µl DMSO and 2 µl Milli-Q® water. The primers enable the amplification of the full-length mutant gene.

Table 2.5: Thermocycler program for overlapping PCR. [1] Extension times were decided by the length of the amplicon using 30 secs/kb. [2] The annealing temperature was determined by the lowest melting temperature of a primer pair and the extension time by the length of the amplified fragment.

Stage	Cycles	Temperature (°C)	Duration
Amplification	30-35	92	30 secs
		50	30-45 secs
		72	[1]
Pause	--	--	--
Denaturation	1	94	3 mins
Amplification	30	94	30 secs
		[2]	15 secs
		72	[1]
Extension	1	72	10 mins
Cooling	--	4	∞

2.1.11.2 QuikChange Site Directed Mutagenesis

To introduce mutations into whole plasmids, the Quikchange Site Directed Mutagenesis method was used. Two primers were designed to have a tail of ~10 nt at the 5' end followed by the mutation and then a clamp of ~30 nt. This created a ~20 nt overlap with the mutation in the centre. PCR reactions were setup as shown in chapter 2.1.2, however the elongation rate was 1 min/kb with only 12 cycles of amplification and an annealing temperature of 60 °C. Controls didn't contain any Phusion Hot Start II DNA Polymerase. After these cycles, 0.5 µl FastDigest DpnI was added and the reaction was left at 37 °C for 60 mins. From this reaction, 10 µl was transformed into chemically competent *E. coli* Bio*Blue*.

2.1.12 DNA Quantification

DNA was quantified either by using the NanoVue Plus spectrophotometer (Version 2.0.3, Biochrom) or the NanoDrop 2000 (Version 1.4.2). The quantity of DNA was calculated by measuring the absorbance of a 2 µl sample at 260 nm. The purity of samples was assessed by the ratio of the 260 nm and the 280 nm absorbance values,

where a ratio of ~1.8 suggested a pure sample. Each run was calibrated with 2 µl of the same Milli-Q® water used to elute the DNA from the purification columns.

2.1.13 Sequencing of DNA

Sequencing reactions were sent to GATC Biotech (European Custom Sequencing, Germany), where results were analysed using the GATCViewer™ 1.00 (GATC Biotech) or SnapGene® Viewer 4.0.7 and the online sequence alignments software, ClustalW2 (EMBL-EBI) or T-coffee (Expasy, Swiss Institute of Bioinformatics). Reference sequences were obtained from National Center for Biotechnology Information (NCBI). For each sample to be sequenced, two reactions were set up; one with the appropriate forward primer and the other with the reverse primer. Sequencing reactions were made up in sterile 1.5 ml microcentrifuge tubes consisting of 5 µl plasmid DNA (80-100 ng/µl), 5 µl forward or reverse primer (10 µM) and made up to 15 µl with Milli-Q® water.

2.1.14 Gene Synthesis

Genes were codon optimised for *E. coli* while *roseRS_3509* and *rcas_0662* were codon optimised for *Parageobacillus thermoglucosidasius* using the online GeneArt® software (Thermo Fisher Scientific). The gene sequences encoding *roseRS_3509*, *rcas_0662*, r*xyl_0493*, *tcur_3107* and *tbis_3257* were synthesized by GeneArt and subcloned into the pMA vector. The gene sequences encoding *tchrom*, *fjsc11*, and *jkg1* were synthesized as blunt DNA sequences by Twist Bioscience (USA). All *E. coli* optimised genes contained a NheI restriction site at the N-termini and an EcoRI restriction site at the C-termini for insertion into the pET28a plasmid except the *Parageobacillus* optimised *roseRS_3509* which contained BsaI restriction sites at both terminals. Codon optimised sequences are shown in Appendix Figure 1.

2.2 Cell Culture

2.2.1 Media

All media was sterilised by autoclaving. For agar plates, 15 g per litre (1.5%) of agar was added to the media prior to autoclaving. The antibiotics, Amp, Kan, Chl and streptomycin, were added to cultures during preparation or to molten agar before pouring.

2.2.1.1 Luria Broth (LB)

1 L of LB media contained 10 g tryptone (1.0%), 5 g yeast extract (0.5%) and 10 g NaCl (1.0%) in Milli-Q® water.

2.2.1.2 SOC Media

Super Optimal broth with Catabolite repression (SOC) media contained 20 g tryptone (2%), 5 g yeast extract (0.5%), 4 g glucose, 0.58 g NaCl, 0.2 g KCl, 2.04 g $MgCl_2.6H_2O$, 1.2 g $MgSO_4$ in 1 L of Milli-Q® water.

2.2.1.3 2XYT Media

2X Tryptone Yeast Broth (2XTY or 2YT) media consisted of 16 g tryptone (1.6%), 10 g yeast extract (1.0%) and 5 g NaCl (0.5%) in 1 L of Milli-Q® water.

2.3 Bacterial Strains

Table 2.6: Bacterial strains used in this study.

Bacterial Strain	Application	Genotype	Supplier
E. coli Bio*Blue*	Plasmid amplification and propagation	*rec*A1, *end*A1 *gyr*A96 *thi*-1 *hsd*R17(r_k-, m_k+) *sup*E44 *rel*A1 *lac* [F' *pro*AB *lac*I^qZΔM15 *Tn*10(Tet^r)]	Bioline, UK
E. coli BL21(DE3)	Protein Expression	F- *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Novagen, Germany
E. coli OverExpress C43(DE3)	Protein Expression	F- *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Lucigen, USA
E. coli S17.1	Plasmid Conjugation into *Parag*eobacillus	F^- *recA1 endA1 thiE1 pro-82 creC510 hsdR17-M+Rp4: 2-Tc:Mu: Km Tn7, λpir lysogen, Sm^R*	
Parageobacillus thermoglucosidasius NCIMB 11955	Strain for terpene production	Wild-type strain	NCIMB Ltd, UK
Parageobacillus thermoglucosidasius NCIMB 11955 cMev9	Strain for terpene production	Wild-type strain with Mev9 with Kan^R integrated into genome at attB site	MQS

2.3.1 Chemically Competent *E. coli*

The Untergasser protocol was used to produce chemically competent *E. coli*. A 10 ml LB culture was inoculated with a loop of the commercial stock of the *E. coli* strain and grown overnight at 37 ºC, 200 rpm. For *E. coli* S17.1, 100 µg/ml streptomycin was added. The overnight culture was used to inoculate two 50 ml cultures of LB media containing 20 mM $MgSO_4$. Once the cultures reached an OD_{600} of between 0.4-0.6, the cultures were pelleted by centrifugation (4000 rpm, 10 mins, 4 ºC). The supernatant was removed and the pellet resuspended gently in 20 ml fresh Inoue

solution (50 mM $MnCl_2$, 20 mM $CaCl_2$, 250 mM KCl, 0.01 M PIPES, pH 6.7). The resuspensions were pooled and pelleted again and the supernatant discarded. The pellet was resuspended in 3 ml Inoue solution with 206.25 µl DMSO added after resuspension. The resuspension was incubated on ice for 10 mins. The cells were aliquoted in 50 µl or 100 µl, into pre-chilled 1.5 ml microcentrifuge tubes and snap frozen in liquid nitrogen. Aliquots were kept at -80 ^{o}C. The full Untergasser protocol is shown at http://www.untergasser.de/lab/protocols/competent_cells_chemical_v1_0.htm.

2.3.2 Transformation

Aliquots of chemically competent *E. coli* were thawed on ice. Pure plasmid DNA, ligation, Quikchange or Golden Gate reactions (1 µl/ 50 µl cells) were added to aliquots of chemically competent *E. coli* and left on ice for 5 mins. For Gibson Assembly, 2.5 µl of each reaction was added to 50 µl chemically competent *E. coli*. Cells were heat-shocked at 42^{o}C for 30 secs before replacing reactions immediately back on ice for a minimum of 1 min. Transformation reactions were recovered in 250 µl of SOC medium for between 45 mins – 2 hrs at 37 ^{o}C, 200 rpm before plating on LB agar plates containing the correct concentration of antibiotic(s).

2.3.3 Glycerol Stocks

Glycerol stocks were made up of 0.5 ml 60% glycerol with 1.0 ml *E. coli*, taken from 10 ml overnight cultures of *E. coli* constructs in LB media. For *Parageobacillus* constructs, 10 ml overnight cultures were pelleted by centrifugation (4000 rpm, 20 ^{o}C, 10 mins) and resuspended in 1.0 ml 2XTY media. The resuspension was added to 0.5 ml 60% glycerol. All stocks were set up in 2.0 ml cryovials. These were stored at -80^{o}C.

2.3.4 Conjugation

Conjugation required overnight cultures of both *P. thermoglucosidasius* NCIMB 11955 or *P. thermoglucosidasius* NCIMB 11955 cMev9 and *E. coli* S17.1, containing

an oriT plasmid, which were set up using either a glycerol stock or a colony from an agar plate. Each culture contained the appropriate antibiotics. If the OD_{600} of the *P. thermoglucosidasius* NCIMB 11955 reached >4 or the *E. coli* S17.1 culture OD_{600} was too low, the *P. thermoglucosidasius* culture needed to be subcultured to an OD_{600} of 0.1. The *P. thermoglucosidasius* NCIMB 11955 culture was cooled to 37°C. To remove the antibiotic from the *E. coli* S17.1 culture, it was pelleted (4000 rpm, 10 mins, room temperature), the supernatant was removed and the cell pellet resuspended in fresh LB media. The same process was performed for the *P. thermoglucosidasius* culture but the cell pellet was resuspended in 2YT media. Using the OD_{600} values, the *E. coli* S17.1 and *P. thermoglucosidasius* NCIMB 11955 were mixed at a ratio of 1:10 with 0.5 ml *E. coli* S17.1 culture combined with the appropriate amount of *P. thermoglucosidasius* NCIMB 11955 culture. Mixtures were pelleted (4000 rpm, 10 mins, room temperature) and the supernatant discarded before the cells were resuspended in the residual media. This was spotted on a LB agar plate containing 10 mM $MgCl_2$ and left at 37°C overnight before being scraped off into 2 ml fresh 2YT media. A dilution series from undiluted to 10^{-5} was set up and three 20 µl spots per dilution were put onto 2YT agar plates. These were incubated upright until dry at 52 °C. Cultures of *P. thermoglucosidasius* NCIMB 11955 used 12.5 µg/ml Kan if it had the chromosomally integrated Mev9 pathway and 5 µg/ml Chl after conjugation.

2.4 Recombinant Protein Expression

2.4.1 *E. coli* Culture

The *E. coli* strains, BL21(DE3) and C43(DE3) were used for recombinant protein expression using the T7lac system.

2.4.1.1 Small-scale

For small scale protein expression of TPSs, 10 ml LB media was inoculated with either a fresh colony or glycerol stock and incubated overnight at room temperature and 40 rpm rocking. LB media cultures (120 ml) containing the appropriate antibiotics

were inoculated with 1 ml of the overnight culture. All cultures contained 50 µg/ml kanamycin. The 120 ml culture was grown to an OD_{600} of 0.6-0.8 before separating it into ten 10 ml cultures. These cultures were used to identify optimum conditions for expression of each TPS. To optimise the IPTG induction concentration, cultures were induced with either 0 µM, 100 µM, 200 µM, 400 µM or 1000 µM IPTG. These cultures were either incubated at 18 ^{o}C or 37 ^{o}C. Samples (1 ml) were taken at 0, 2, 4, 21 and 24 hpi at 18 ^{o}C and 25 ^{o}C and 0, 30 mins, 1, 2 and 3 hpi at 37 ^{o}C. Cells were pelleted (13000 rpm, 5 mins) and the medium discarded. Pellets were stored at -20 ^{o}C until the samples could be analysed by SDS-PAGE.

The optimised temperature for protein production was 18 ºC for all TPSs produced. The IPTG concentrations for induction of Tcur_3107 and FJSC11 production was 100 µM, 200 µM for Tchrom and Tbis_3257 and 400 µM for JKG1. All cultures were pelleted after 21 hrs except Tchrom and FJSC11 which were harvested after 24 hrs. The optimal production conditions for SSCG_03688 were taken from Hu *et al.* (2011), for RoseRS_3509 and Rcas_0662 were taken from Styles *et al.* (2017) and for selinadiene synthase (SdS) was taken from *Baer et al.* (2014). The same conditions for RoseRS_3509 was used for Rxyl_0493.

2.4.1.2 Large-scale

Overnight cultures were prepared as for small scale (See above), however, 1 ml of the overnight culture was used to inoculate 500 ml LB media. The total volume of LB media inoculated for each protein expression was between 1 and 8 L. At an OD_{600} of ~0.8, the cultures were induced with IPTG and incubated using the predetermined optimum expression conditions. Cultures were pelleted (7000 rpm, 10 mins) and stored at -20 ^{o}C.

2.4.2 Cell lysis

Small scale culture samples (1 ml) were lysed by freeze-thawing. Cell pellets were resuspended in 100 µl resuspension buffer and freeze-thawed three times by incubating at -80 ^{o}C for 5 mins and 42 ^{o}C for 10 mins. The resuspension buffer for Tchrom and FJSC11 was 20 mM Tris, pH 8.0, 300 mM NaCl, 10% glycerol, 20 mM imidazole. The resuspension buffer (10 ml) for Tbis_3257 included 1 mg/ml lysozyme

and a protease inhibitor cocktail tablet (Roche, Switzerland) while the resuspension buffer for JKG1 was at pH 7.0. The resuspension buffer for Tcur_3107 was 20 mM Tris, pH 8.0, 300 mM NaCl and 1 mg/ml lysozyme. All DNA was removed by adding 1 µl DNase (1 U/µl) and 10 µl 10x Reaction Buffer with $MgCl_2$ and incubating at room temperature (10 mins). To separate soluble and insoluble fractions, samples were centrifuged (13000 rpm, 10 mins) and the soluble fraction transferred to a fresh microcentrifuge tube. Only the insoluble fractions were resuspended for Tbis_3257 in 100 µl of resuspension buffer. Samples were analysed by SDS-PAGE gel electrophoresis.

Large scale culture pellets were resuspended in 10-15 ml resuspension buffer per 500 ml of bacterial culture. All resuspension buffers contained 1 mg/ml lysozyme and a protease inhibitor cocktail tablet for every 40 ml of resuspension buffer. The resuspension buffer for SSCG_03688 and all of its truncation mutants was 50 mM Tris, pH 7.5, 500 mM NaCl, 20 mM imidazole, and 5 mM β-mercaptoethanol according to Hu *et al.* (2011). The resuspension buffer for further SSCG_R6 mutants cell pellets included 10 mM $MgCl_2$. The resuspension buffer for SdS was as in Baer *et al.* (2014). The resuspension buffer for Tchrom and FJSC11 contained 20 mM Tris, pH 8.0, 10% glycerol, 300 mM NaCl, 10 mM $MgCl_2$, 20 mM imidazole. The same resuspension buffer was used for RoseRS_3509 and Rcas_0662 being purified for crystallisation. RoseRS_3509 and the active site mutants used the same resuspension buffer but at pH 8.5. The resuspension buffer for JKG1 was identical except it was at pH 7.0 and for Tcur_3107 included 5 mM β-mercaptoethanol and excluded $MgCl_2$. Each resuspension was sonicated (Soniprep 150 Plus with the exponential probe, Measuring and Scientific Equipment Ltd, UK) at amplitude 14 for 15 secs between 3-5 times while on ice. Each sample was placed on ice and mixed between sonications. The cell lysates were separated according to solubility by centrifuging twice at 10000 rpm for 10 mins with the soluble supernatant being transferred to a separate sterile tube each time. The soluble supernatants were purified by affinity chromatography.

2.4.3 Affinity Chromatography

The soluble fraction of the cell lysate was loaded onto 0.5-1.0 ml TALON Metal Affinity Resin (Clontech, USA) in a 50 ml sterile falcon tube. Prior to this, the resin was equilibrated with 20 column volumes (CVs) of Milli-Q® water and Resuspension Buffer

(Without lysozyme, protease inhibitor cocktail tablet, β-mercaptoethanol or DTT if used) with a centrifugation step (4000 rpm, 10 mins, 4 ^{o}C) between each step to pellet the resin and pour off the supernatant. The resin was resuspended in the soluble cell lysate to allow for batch binding (60 rpm, 4 ^{o}C, 30 mins). The column was centrifuged to pellet the resin and the supernatant poured into a fresh 50 ml falcon tube and labelled the flow-through (FT). The resin was resuspended in 10-20 CVs wash buffer (WB) before rolling again for 10 mins. The resin was pelleted again and most of the wash buffer removed. The remaining WB was used to resuspend the resin and transfer it to the polypropylene column where the wash buffer flowed through. Alternatively, the resin was placed into the polypropylene column before adding the WB. For purification of FJSC11, Tbis_3257 and Rxyl_0493, two wash buffer steps were performed. The FT and WB solutions were kept at 4 ^{o}C. To elute the protein, 2 CVs of elution buffer was added and the elution (E) collected. The elution step was performed 1-3 times. Samples of soluble and insoluble cell lysate, FT, WB and E were run on a SDS-PAGE gel to determine protein solubility. Wash buffer and elution buffer compositions are shown in Table 2.7.

Table 2.7: Difference in wash buffer and elution buffer compositions between each TPS purification. The wash buffer used for all TPS purifications contained 20 mM Tris, pH 8.0, 10 % glycerol, 300 mM NaCl, 10 mM $MgCl_2$, 50 mM imidazole and the elution buffer had the same composition except it contained 500 mM imidazole. Any differences between or additions to the buffers used for purification with specific TPSs are shown below.

Protein	**Wash Buffer**	**Elution Buffer**
Tcur_3107	5 mM β-mercaptoethanol	5 mM β-mercaptoethanol
SSCG_03688 and truncation mutants	pH 7.5, 5 mM β-mercaptoethanol	pH 7.5, 5 mM β-mercaptoethanol
SdS and mutants	50 mM NaCl (No $MgCl_2$ or imidazole)	500 mM NaCl, 400 mM imidazole
RoseRS_3509 and product distribution mutants	pH 8.5	pH 8.5

2.4.4 Size Exclusion Chromatography

A Superdex™ 200 prep grade (HiLoad™ 16/60, GE Healthcare Life Science, UK; column volume, 120 ml) was equilibrated using Milli-Q® water followed by size exclusion buffer between 0.2-0.5 ml/min including the flushing of the pumps on an

ÄKTA FPLC (GE Healthcare Life Science, UK). The size exclusion buffer for Rcas_0662 contained 20 mM Tris, pH 8.0, 10% glycerol, 10 mM $MgCl_2$, 100 mM NaCl. For RoseRS_3509, the size exclusion buffer was also at pH 7.5 and contained 5 mM DTT. The soluble elution fractions from the affinity chromatography were combined and loaded onto the column using a 50 ml superloop or 2 ml sample loop (GE Healthcare Life Science, UK) and the flow rate switched to 1.0 ml/min of SE buffer. Fractions (1 ml) were collected after 40 mins using a fraction collector. The UNICORN software (Version 5, GE Healthcare Life Sciences, UK) used to run the ÄKTA FPLC allowed the column progress to be followed including the 280 nm absorbance for protein elution. This enabled fraction selection for SDS-PAGE gel electrophoresis, fraction pooling and protein quantification of the purified TPS.

2.4.5 SDS-PAGE Gel Electrophoresis

The insoluble cell debris pellets from small scale cultures were resuspended in 50 µl Milli-Q® while large scale insoluble cell pellets had a loop resuspended in 1.0 ml Milli-Q® water. The insoluble material of the FT sample (1 ml), WB sample (1 ml) and elutions was separated by centrifugation (13000 rpm, 10 mins). The soluble fractions were transferred to fresh microcentrifuge tubes and the insoluble pellets resuspended in 0.10-0.25 ml resuspension buffer. From each sample, 20 µl was placed in a 0.2 ml PCR tube and the samples were denatured by adding 5x protein loading buffer and heat treated at 98 ^{o}C for 10 mins before loading onto a 12 % SDS Precast gel (Expedeon, UK). To determine protein size, Unstained Protein MW Marker was used as a molecular marker (Figure 2.2). The reservoir of a Mini-PROTEAN® Tetra Cell (Bio-Rad, UK) was filled with 1x SDS Run Buffer: Teo-Tricine (Expedeon, UK) and resolved at 70 mA/gel for 35-60 mins using a PowerPac™ Basic (Bio-Rad, UK). Gels were removed from their cast and stained with Coomassie Blue (3 mM Coomassie Brilliant Blue, 30% methanol and 10% acetic acid). When the gels were sufficiently stained, the Coomassie Blue solution was replaced by SDS-PAGE gel destain (30% methanol and 10% acetic acid) after washing in dH_2O. Once sufficiently destained, images were taken with the G:Box (Syngene, USA) using Gene Snap software (Version 7.09.17, Syngene, USA). In the case of size exclusion, the SDS-PAGE gels were used to identify which fractions to combine.

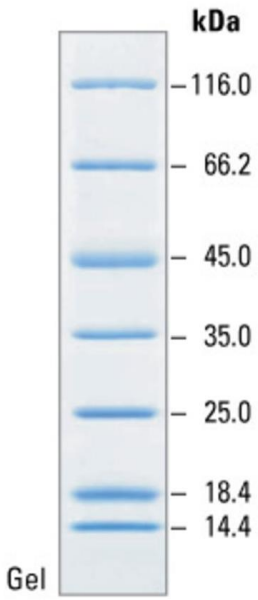


Figure 2.2: Unstained Protein Molecular Weight Marker used for estimating the size of protein bands on SDS-PAGE gels.

2.4.6 Protein Concentration

Amicon® Ultra – 15 Centrifugal Filters 10 kDA MWCO and 30 kDa MWCO tubes (Merck Millipore, Ireland) were used to concentrate and exchange buffer of purified protein solutions for storage. After affinity chromatography, proteins were concentrated to >3 mg/ml and buffer exchanged to reduce the imidazole concentration to <1 mM through three rounds of buffer exchange (7000 rpm, 4 ºC). The exchange buffer for each TPS only purified by affinity chromatography is shown in Table 2.8. After size exclusion chromatography, samples were concentrated to >7 mg/ml for protein crystallisation by centrifugation (7000 rpm, 4 ºC). The tube membrane was washed five times to remove any residual protein on the membranes before the protein sample was transferred to a sterile tube. Concentrated protein samples were separated into 30-50 µl aliquots before being flash frozen with liquid nitrogen. Aliquots were stored at -80 ºC.

Table 2.8: Differences in exchange buffer composition between each TPS purification. The exchange buffer contained 20 mM Tris, pH 7.5, 10 % glycerol, 300 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT. Any differences or additions to the exchange buffer used for specific TPS purification are shown in the table below.

Protein	Exchange Buffer
JKG1	pH 6.5
Tcur_3107	pH 8.0, 100 mM NaCl, 5 mM β-mercaptoethanol instead of DTT
SSCG_03688 and truncation mutants	50 mM PIPES instead of Tris, 100 mM NaCl, 20 % glycerol, 5 mM β-mercaptoethanol instead of DTT
SSCG_R6 and thermostability mutants	100 mM NaCl, 5 mM β-mercaptoethanol instead of DTT
SdS and thermostability mutants	100 mM NaCl
RoseRS_3509 and product distribution mutants	pH 8.5, 100 mM NaCl, no DTT

2.4.7 Protein Quantification

All purified protein was quantified using the Bradford Protein Assay kit (Bio-Rad, UK). For the cuvette method, a 2 mg/ml BSA standard was diluted to 100 µg/ml in the same buffer as the TPS. This was further diluted in the same buffer in 100 µl with 900 µl of Bio-RAD Protein Assay Dye Reagent (Bio-Rad, UK) added to create a range of BSA concentrations between 0-10 µg/ml. The OD_{595} value of each standard enabled a calibration curve to be plotted. Purified protein samples were prepared similarly to the BSA standards and the calibration curve was used to identify the concentration of the purified protein. This protocol was adapted to using 96 microtiter plates (Griener BioOne, UK) also. BSA standards between 0-10 µg/ml were made up in 1 ml volumes. The Bio-RAD Protein Assay Dye Reagent was diluted 2:7 in Milli-Q® water. In the plate, each well contained 100 µl of a BSA standard and 100 µl of the diluted Bio-RAD Protein Assay Dye Reagent. Samples were left at room temperature for 5 mins before the OD_{595} readings were taken using a BioTek Synergy HT plate reader (BioTek, UK) for the calibration curve. TPS samples were prepared in the same way as the BSA standards. All standards were set up in duplicate and the TPS samples in triplicate.

2.5 Hidden Markov Model

Hidden Markov models were run using the online software, HMMER3 (EMBL-EBI, UK), where sequence alignments were inserted into the HMMsearch algorithm (Finn *et al.* 2015). Alignments were created in the aligned FASTA format using the multiple sequence alignment tool, T-Coffee (EMBL-EBI, UK). Taxonomic restrictions were applied to only identify bacterial sequences. The sequence and hit E-value score cut-offs set to 1×10^{-5}. The HMM was performed against the UniProtKB sequence database.

2.6 Protein Study

2.6.1 Protein Model Design

Initial TPS models were created using the online software tool, SWISS-MODEL (Version 8.05) (Waterhouse *et al.* 2018). Through SWISS-MODEL, 3D homology models were built from the high-resolution crystallised proteins in the PDB database (https://www.rcsb.org) with the highest sequence homology to the query. For more accurate homology models, I-TASSER was used (Yang and Zhang 2015). I-TASSER builds a homology model using similar folds from crystal structures in the PDB database before using various algorithms to assemble a complete structure. I-TASSER provided multiple models where the one chosen for further analysis had the highest C-score and lowest root mean square deviation (RMSD). Protein model files were exported to PyMol software (Version 1.7.4.5, Schrödinger, LLC., USA) for visualisation of the homology model, predicted active site cavities and conserved motif positions. Structural alignments in PyMol were also performed using the cealign command.

2.6.2 Crystallisation

To determine the concentration for crystallising the TPSs, the pre-crystallisation test (PCT, Hampton Research, USA) was used and followed as per the manufacturer's instructions. Rather than using the kit itself, all of the buffers were made in-house.

Using the concentration determined by the PCT, crystallisation condition screening was performed for RoseRS_3509 and Rcas_0662 with a range of pre-prepared HT-96-well screens (Molecular Dimensions, UK) using a Crystal Phoenix nano dispensing robot (Art Robbins Instruments, LLC, USA). All plates used the sitting drop method with three 0.3 µl drops of 1:2, 1:1 and 2:1 ratios of precipitant and protein. Plates were checked for crystals using a wild M5A microscope (Wild Heerbrugg, Switzerland). The HT-96 Molecular Dimensions screens used were:

- JCSG-plus™ HT-96
- SG1™ (Shotgun) HT-96
- CSSI (Clear Strategy™ Screen I) HT-96
- CSSII (Clear Strategy™ Screen II) HT-96
- SSI & II (Structure Screen 1 + 2) HT-96
- PACT premier™ HT-96
- MIDAS™ HT-96
- ProPlex HT-96

2.6.3 GC-MS

The organic layer (100 µl) of all enzyme assays was placed into a 200 µl screw top vial with an insert (Agilent Technologies, UK) to be analysed by the Agilent 7890B GC-MS system (Agilent Technologies, UK) to determine the terpene products qualitatively and quantitatively. To separate samples, a DB-FFAP capillary GC column (30 m x 250 µm x 0.25 µm, Agilent Technologies, UK) was used. 1 µl of sample was injected onto the column. For 1 min, the column was held at 40 ^{o}C before the temperature was ramped up to 250 ^{o}C at a rate of 20 ^{o}C/min, which was held for a further 8 mins. All terpene compounds typically eluted between 8 mins and 15 mins, and the total method time was 19.5 mins. To quantify the amount of terpene product present in each sample, a β-caryophyllene internal standard or calibration curve was used and the areas of the peaks from the FID (Flame ionisation detector) were compared. For the identification of unknown products, authentic standards were used where possible to confirm the terpene product. Retention times of terpene products varied between different runs as the column length was shortened with age. 1 mg/ml stock solutions of GPP (2.7 mM, Sigma-Aldrich, UK) and FPP (2.3 mM, Isoprenoids, USA) were used and stored at -20 ^{o}C. All prenyl diphosphates and their dilutions were stored in 7:3 methanol:dH_2O with 10 mM NH_4OH.

2.6.3.1 Activity Assay

The vial assay developed by Garrett *et al* (2012) was adapted for this study. Qualitative assays were performed for Tchrom, Tbis_3257, JKG1 and FJSC11 to identify whether they had any activity and with which prenyl diphosphate, GPP or FPP. In 200 µl reactions, 4 µl of either GPP or FPP were added to 192 µl assay buffer (25 mM MES, 25 mM CAPS, 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM $MgCl_2$) and the 0.5 ml microcentrifuge tubes warmed to 50 ºC. From 50 µM stocks of TPS, 4 µl was added to the reaction and it was mixed before being left at 50 ºC for 1 hour. The final concentrations were 1 µM, 54.8 µM and 46.2 µM for the TPS, GPP and FPP, respectively. A hexane overlay of 150 µl without an internal standard was added for extraction. The qualitative assay for Tcur_3107 was performed at 37 ºC rather than 50 ºC. This contained 160 µl assay buffer (25 mM MES, 25 mM CAPS, 50 mM Tris pH 7.0, 300 mM NaCl, 10 mM $MgCl_2$) with 20 µl GPP or 8 µl FPP. The GPP was a concentrated stock of > 1 mg/ml provided by Marc Pfleiger (University of Bath, UK). This was warmed at 37 ºC before the addition of 5 µl of 50 µM Tcur_3107. After the reaction was started, a 150 µl hexane overlay containing 10 µg/ml caryophyllene was added. Reactions were terminated by freezing in liquid nitrogen before thawing by vortexing. Any precipitation was pelleted (13000 rpm, 5 min) before taking 100 µl of the hexane overlayer for GC-MS analysis. All reactions were performed in triplicate.

As in Hu *et al.* (2011), qualitative assays of SSCG_03688 and SSCG_R6 at 30 ºC and 40 ºC were performed in an assay buffer containing 50 mM PIPES, pH 6.7, 100 mM NaCl, 15 mM $MgCl_2$, 20 % glycerol, 5 mM β-mercaptoethanol. Each 100 µl reaction contained 1 µM SSCG_03688 or SSCG_R6 and 92.3 µM FPP. The reactions were incubated at the appropriate temperature before flash freezing in liquid nitrogen. The hexane overlayer, containing 10 µg/ml caryophyllene (150 µl), was added with the organic and aqueous phases being mixed by vortexing. After separating the phases by centrifugation (1000 rpm, 1 min), 100 µl of the organic phase was taken for GC-MS analysis.

2.6.3.2 Kinetics Assay

Reactions had varying concentrations of FPP between 1.5 µM and 115.4 µM with an enzyme concentration of 1 µM in 200 µl. The reactions were snap frozen in liquid nitrogen to terminate the reaction. On removal from the liquid nitrogen, the organic

layer was added to the reactions containing 10 µg/ml caryophyllene. The reaction was vortexed until it was completely thawed. The organic phase was extracted into a 250 µl GC-MS vial. For SSCG_03688 and SSCG_R6 mutant, all kinetic assays were performed in triplicate at 30 °C and reactions were terminated after 5 mins. To quantify the amount of terpene in each sample, GC-FID was used. As all enzymes were shown to follow Michaelis-Menten single substrate kinetics, the data were fit to the Michaelis-Menten equation (Equation 1) using SigmaPlot software (Version 12.3, Systat Software, San Jose, CA) to calculate its various parameters.

Equation 1: $$v = \frac{V_{max}[S]}{K_M + [S]}$$

2.6.3.3 Irreversible Denaturation Assay

Initial reactions were set up in 500 µl microcentrifuge tubes. These tubes contained a 12.5 µM stock of SSCG_03688 in assay buffer (25 mM MES, 25 mM CAPS, 50 mM Tris, 10 mM $MgCl_2$, 300 mM NaCl). Single replicates were heated to 35 ºC, 40 ºC, 45 ºC and 50 ºC with samples (4 µl) taken for immediate activity assay when the inactivation assay was initiated and then at ten-minute time points for an hour. These samples were added to a 500 µl microcentrifuge tube containing 43 µl assay buffer kept at 30 ºC. 3 µl FPP (2.3 mM) was added to initiate the assay which was terminated by placing the tube into liquid nitrogen. Reactions were extracted by adding 150 µl hexane and thawing the reactions by vortexing to mix the organic and aqueous phases. The phases were separated by centrifuging at 2000 rpm for 1 min. From the organic top layer, 100 µl was analysed by GC-FID analysis. For SSCG_03688 related enzymes, 35 ºC was chosen as the denaturation temperature. Assays of SSCG_03688, SSCG_R6, SSCG_R6_G24I, SSCG_R6_S100Q and SSCG_R6_G209A were run in triplicate. Denaturation assays were performed for SdS but the data was unreliable. The rate of inactivation was calculated using equations 2 and 3.

Equation 2: $\frac{c}{c_o} = e^{-kt}$ *Equation 3:* $Ln\left(\frac{c}{c_o} \times 100\right) = -kt$

In this formula, c represents the concentration of active and reversibly inactive protein at a specific time-point with c_o representing the concentration of active and reversibly inactive protein at time-point zero. The fold change between c and c_o plotted against time, t, can be used to calculate the inactivation rate, k, from the negative gradient.

The fold change between c and c_o was converted to a percentage in this case (Equation 3).

2.6.4 Malachite Green Assay

Kinetic assays were performed on Tchrom, SSCG_R6 mutants, SdS mutants, and RoseRS_3509 and its product distribution mutants using the malachite green assay as in Vardakou *et al.* (2014). The substrate concentrations were between 0-115.4 µM for FPP and 0-92.4 µM for GPP. The TPS concentration was 1 µM except for RoseRS_3509 which was 0.2 µM. Assays were performed at 50 ^oC for RoseRS_3509 and its mutants (W296A, F78A and Y310A) and 30 ^oC for all SSCG_R6 and SdS mutants. The assay buffer (25 mM MES, 25 mM CAPS, 50 mM Tris, 10 mM $MgCl_2$, 300 mM NaCl, 2 U/µl inorganic pyrophosphatase) was pre-warmed before each assay for 2 mins prior to addition of enzyme and substrate. The pH of the assay buffer was pH 7.0 for SdS and its mutants, pH 7.5 for SSCG_R6 and its mutants and pH 8.0 for Tchrom, RoseRS_3509 and its production mutants. Each assay was performed in triplicate. After the TPS was added, assays were run for 5 mins before quenching with the malachite green development solution (2.6 mM malachite green, 2.4 M sulfuric acid, 1.48 % ammonium molybdate and 0.16 % Tween20). This mixture was left at room temperature and shaking at 200 rpm. Background assays without the TPS present were run at the same time. All assays were performed in triplicate. Each sample was read on a 96-microtiter plate (Greiner BioOne, UK) at 623 nm using a plate reader. The kinetic parameters were calculated using Equation 1. The turnover rate of yeast inorganic pyrophosphatase used in this assay was shown to much higher than any TPS by Zyryanov et al (2002).

2.6.5 Circular Dichroism

Each TPS was diluted to 1 µM in CD buffer (100 mM KF, 10 mM sodium phosphate) in a total volume of 250 µl and placed into a 1 mm path length quartz cuvette. Wavelength scans between 200-300 nm were performed by a Chirascan CD Spectrometer (Applied Photophysics Ltd, UK). Protein wavelength scans were performed between 200-300 nm at 20 ^oC to determine the protein secondary structure of the TPSs. Protein thermal melts were performed at 222 nm with the temperature

increasing between 10-95 °C (TPS dependent) in 1 °C increments, where each temperature was held for 5 secs. To determine the T_m of each TPS, the two-state transition equation (Equation 4) was used as described by Catici *et al.* (2015). The slope and intercept of the unfolded (u) and folded (f) baselines are b and m respectively. ΔH is the van't Hoff enthalpy of unfolding at T_m. The data of the melt curves was fitted to equation 4 using the program, SciDavis (http://scidavis.sourceforge.net/).

Equation 4: $$\theta_{222\,nm} = \frac{b_u + m_u T + (b_f + m_f T) K_u}{1 + K_u}$$

Equation 5: $$K_u = \exp\left(\frac{\Delta H\left(1 - \frac{T}{Tm}\right)}{RT}\right)$$

2.7 2,3-Dihydrofarnesyl diphosphate synthesis

The chemical synthesis of 2,3-dihydrofarnesyl diphosphate (DHFPP) from farnesol was performed by Robert Chapman (Department of Chemistry, University of Bath). The first step (Figure 2.3A) was performed according to Suhara *et al.* (2010) converting farnesol to (E)-3,7,11-trimethyldodeca-6,10-dien-1-ol (4.5 g, 89%). ^{1}H NMR (500 MHz, $CDCl_3$ δ 0.91 (3H, d, J = 6.5 Hz), 1.18 - 1.43 (5H, m), 1.56 - 1.65 (1H, m), 1.60 (6H, s), 1.68 (3H, s), 1.95 - 2.09 (6H, m), 3.64 - 3.73 (2H, m), 5.07 - 5.12 (2H, m); ^{13}C NMR (125 MHz, $CDCl_3$) δ 15.9,17.7, 19.5, 25.3, 26.7, 29.2, 37.2, 39.7, 39.9, 61.2, 124.4, 124.6, 131.3, 134.9.

The second step (Figure 2.3B) also followed Suhara *et al.* (2010) but with changes in the amounts of the reagents added. The initial solution of (E)-3,7,11-trimethyldodeca-6,10-dien-1-ol (4.00 g, 17.8 mmol) was combined with dichloromethane (50 ml), methanesulfonyl chloride (2.40 g, 21.0 mmol) and triethylamine (3.24g, 44.0 mmol). The residue produced after washing, filtering and concentrating was dissolved in acetone (40 ml) and sodium iodide (10.0 g, 66.6 mmol) added. The yield of (*E*)-12-iodo-2,6,10-trimethyldodeca-2,6-diene was 4.10 g (69%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.88 (3H, d, J =6.5 Hz), 1.15 - 1.49 (4H, m), 1.56 - 1.65 (1H, m), 1.60 (6H, s), 1.68 (3H, s), 1.85 - 2.09 (6H, m), 3.14-3.27 (2H, m), 5.08 - 5.12 (2H, m); ^{13}C NMR (125 MHz, $CDCl_3$) δ 16.0, 17.7, 18.7, 25.2, 25.7, 26.7, 33.5, 36.3, 39.7, 40.9, 124.30, 124.35, 131.3, 135.1.

The final step (Figure 2.3C) was performed according to the protocols by Baer *et al.* (2014) and Woodside *et al.* (1993) other than different starting quantities of

Tris(tetrabutylammonium)-hydrogen diphosphate (2.43 g, 2.69 mmol), acetonitrile (20 ml) and (*E*)-12-iodo-2,6,10-trimethyldodeca-2,6-diene (0.60 g, 1.79 mmol). The resulting 2,3-dihydrofarnesyl diphosphate was stored at -20 ºC. The ^{1}H-NMR spectrum of (E)-2,3-dihydrofarnesyl diphosphate is on the attached supplementary media under the filename "(E)-2,3-dihydrofarnesyl diphosphate NMR 1H". ^{1}H NMR (300 MHz, Deuterium Oxide) δ 5.16 (app dd, *J*= 16.3, 7.7 Hz, 2H, C=C*H*), 4.01 - 3.79 (m, 2H, CH_2O), 2.01 (dd, *J* = 25.9, 7.1 Hz, 6H, 3x CH_2), 1.78 – 1.63 (m, 1H, C*H*), 1.62 (s, 3H, CH_3), 1.56 (s. 6H, 2x CH_3), 1.48 - 1.02 (m, 4H, 2x CH_2), 0.84 (d, *J* = 6.5 Hz, 3H).

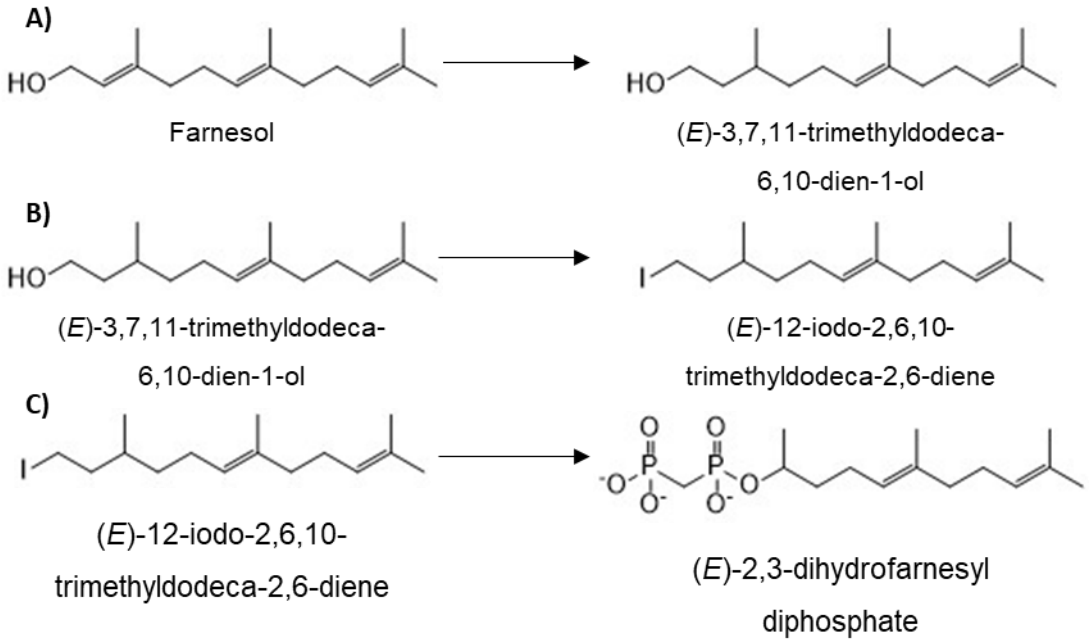


Figure 2.3: Chemical synthesis of (E)-2,3-dihydrofarnesyl diphosphate from farnesol. A) The initial step converts farnesol to (E)-3,7,11-trimethyldodeca-6,10-dien-1-ol. B) (E)-3,7,11-trimethyldodeca-6,10-dien-1-ol is further converted to (E)-12-iodo-2,6,10-trimethyldodeca-2,6-diene. C) The final step is the formation of (E)-2,3-dihydrofarnesyl diphosphate from (E)-12-iodo-2,6,10-trimethyldodeca-2,6-diene.

2.8 DHFPP incubation with RoseRS_3509

On advice from personal communication with Philipp Baer, DHFPP was added directly to RoseRS_3509. A stock of RoseRS_3509 was diluted to 10 mg/ml in SEC buffer and 200 µl added to a freeze-dried aliquot of DHFPP (0.2 mg). Any precipitate was removed by centrifugation (13000 rpm, 10 mins) and the RoseRS_3509

quantified by Bradford assay. The RoseRS_3509-DHFPP mixture was then used on the PCTs before putting the mixture on crystallisation plates.

2.9 DHFPP and FPP Active Site Docking

Computational docking experiments were performed by Rory Crean (University of Bath, UK). Docking of FPP and DHFPP into the active site of the I-TASSER model of RoseRS_3509 was performed using Autodock Vina (Trott and Olson 2010) and the online server, RosettaLigand (DeLuca *et al.* 2015). All input files were in PDB file format. For Autodock Vina, Gasteiger charger were used and the search exhaustiveness set to 32.

3 Bioinformatic search for novel thermostable terpene synthases

3.1 Introduction

The first bacterially produced terpene, geosmin, was reported in 1965 from multiple *Streptomyces* species (Gerber and Lechevalier 1965), but it took another two decades before the associated enzymes for the formation of terpenes, the terpene synthases (TPSs), were isolated and discussed (Croteau and Cane 1985). Since then, the search for new TPSs has unveiled thousands of TPS encoding sequences mainly from the genomes of plant and fungi kingdoms. However, the terpene profile of an organism is normally identified first before identifying the associated TPSs. Plants have diverse terpene profiles and therefore generally encode multiple TPSs, whereas bacteria usually have fewer TPSs, resulting in more research being performed on plant TPSs (Cane *et al.* 1994, Lin *et al.* 2006, Agger *et al.* 2008, Hu *et al.* 2011, Nakano, Kim, *et al.* 2011a). In the last 20 years, as more bacterial genome sequences have become available however, more TPSs have been isolated and characterised from bacteria, which has shown that they can produce a wide range of terpene structures (Yamada *et al.* 2015, Dickschat 2016). As discussed in chapter 1.3, TPSs can be divided into two classes which are determined by domain structure and reaction mechanism. Class I bacterial TPSs contain a single α helical bundle domain. In bioinformatics this is the recognised Pfam domain, Terpene_synth_C or PF03936. All class I TPSs initiate carbocation formation by removal of the diphosphate from the GPP, FPP or GGPP. In contrast, class II bacterial TPSs contain two domains which both form α-barrel domains, called the γ and β domains with the interface of these two domains forming the active site. Plant TPS structures are generally considered more complex than bacterial structures as all plant TPSs contain the β-domain. This domain is recognised in the Pfam database as the N-terminal domain or PF01397. Most plant class I enzymes have either a βα didomain structure or the γβα tridomain structure while class II enzymes tend to only have γβα structures (Kawaide *et al.* 1997, Oikawa *et al.* 2001, Zhou, Gao, Hoy, *et al.* 2012).

Plant TPSs usually share higher protein sequence identities with each other than bacterial TPSs do with one another. The majority of plant TPSs contain a common domain, the β domain (PF01397), with some shared identity between and within class I and II enzymes. However, the β domain is not labelled in every plant TPS as a

PF01397 domain. However, bacterial enzymes do not share a common domain between the classes (Bohlmann *et al.* 1997, 1998, Martin *et al.* 2004, Jia *et al.* 2005, Köksal *et al.* 2011). The level of sequence similarity and identity between the β-domains of plant TPSs means that many can be identified by simple search algorithms such as the Basic Local Alignment Search Tool or BLAST (Trapp and Croteau 2001b, Jia *et al.* 2005, Nieuwenhuizen *et al.* 2009). BLAST searches are able to find homologous proteins with high protein similarity, however, as the sequence identity drops to below 30%, BLAST will identify fewer sequences (Skewes-Cox *et al.* 2014). This is not very useful when searching for bacterial class I TPSs that can have no extended regions of high similarity other than several short, conserved motifs. To enable sequences with lower homologies to be identified, more complex algorithms are required. Some method examples are Position Specific Iterated-BLAST (PSI-BLAST) and the commonly used Hidden Markov Models (HMMs). To begin a PSI-BLAST, a BLAST search using a single sequence is run to identify homologous sequences. Selected homologous sequences can then be used to form a multiple sequence alignment with the initial search sequence, on which further BLAST searches can be performed. Each iteration adds more homologous sequences to the alignment and the searches are continued until no new sequences are identified (Eddy 1998). This makes PSI-BLAST much more sensitive at identifying remote sequences than individual BLAST searches. HMMs are statistical models that use probability to predict the next observation in a sequence of "observable events" through "invisible states", both of which are stochastic processes (Eddy 1998, Yoon 2009). For a protein sequence, the type of amino acid is the observable event and the hidden states are the positions in the sequence. From a sequence alignment, the HMM would predict the probability of transitioning between states, known as a Markov chain, as well as the probability of each amino acid in the next position of the sequence. This probability will vary from state to state (Yoon 2009).

One type of HMM that can be used to find new homologous sequences is called a profile-HMM. Where a regular HMM allows cycles between states, a profile-HMM forces the sequence to move forward linearly from the starting state towards the end state. It also introduces insert and delete states, where insert states allow multiple emissions in between match states and delete states are silent and do not provide an emission. These allow for insertions and deletions of residues between different sequences. By submitting a sequence alignment of related sequences, a profile-HMM can be built with probabilities of each state emission, transition, insertion and deletion. New sequences can then be compared to the model and if they reach a threshold

score will be output as a hit. Certain databases, such as Pfam and PROSITE, use profile-HMMs to identify whether new sequences relate to a specific family, contain a specific domain or fold (Sigrist *et al.* 2002, El-Gebali *et al.* 2018). Profile-HMMs identify regions of high homology such as motifs and score these more highly, therefore more distantly related sequences of proteins with similar functions can be identified. As class I bacterial TPSs can have very low sequence homologies, profile-HMMs have been used to identify novel TPSs among bacteria.

This approach was first used by Komatsu *et al.* (2008) to identify the bacterial genes associated with the biosynthesis of the monoterpene, 2-methylisoborneol. Through this study, 41 genes were identified and predicted to encode a range of mono-, sesqui- and diterpene synthases, six of which were shown to produce 2-methylisoborneol and one to produce 2-methylenebornane. Most of these predicted TPSs have since been characterised including the sesquiterpene synthases, avermitlol and (+)-*epi*-cubenol synthases (Chou *et al.* 2010, Nakano *et al.* 2012, Yamada *et al.* 2012, 2015). Using the newly predicted TPSs, Yamada *et al.* (2012) were able to re-train the initial profile-HMM with the 41 sequences to further predict a total of 140 bacterial TPS sequences. These sequences were used by Yamada *et al.* (2015) again to retrain the profile-HMM once more to give a total of 262 predicted sequences. Phylogenetic analyses revealed that some of these TPSs separated into product specific clades including geosmin, *epi*-isozizaene and 2-methylisoborneol synthases which are common bacterial terpene products. A significant number of the sequences were characterised from *Actinomycetales*, mainly *Streptomyces* species, which have been extensively investigated for their huge diversity of natural products with potential value as pharmaceuticals and in industry (Wang *et al.* 2005, Ikeda *et al.* 2014, Keulen and Dyson 2014, Liu, Deng, *et al.* 2018). With the class I TPS profile-HMM, Yamada et al (2012) also performed a class II diterpene synthase profile-HMM to sample its diversity. Only 38 sequences were identified, and none have been characterised, but over two thirds were also from *Actinomycetales*. This group of profile-HMM studies demonstrated that the bacterial terpenome is much more diverse than originally believed but is still significantly less diverse than the plant terpenome (Christianson 2008, Chen *et al.* 2009, Dickschat 2016, Priya *et al.* 2016, Vattekkatte *et al.* 2018). Of the sequences returned by the profile-HMMs, only a few sequences from thermophilic organisms were identified.

The class I profile-HMM from Yamada *et al.* (2015) contained two thermostable TPSs that have since been characterised. These were the naturally thermostable τ-muurolol synthases, RoseRS_3509 and Rcas_0662 from *Roseiflexus* sp. RS-1 and

Roseiflexus castenholzii DSM 13941, respectively (Styles *et al.* 2017). There were two further sequences that could encode thermostable TPSs as they originate from thermophilic bacteria. These were Tcur_3107 from *Thermomonospora curvata* DSM 43183 and Rxyl_0493 from *Rubrobacter xylanophilus* DSM 9941; the latter has been characterised to encode diterpene synthase which produces, cembrene C. From the class II diterpene synthase profile-HMM, only three sequences from thermophiles were identified; namely *Thermomonospora curvata*, *Thermobispora bispora* and *Saccharomonospora viridis*. HMMs are now also being used to identify more diversity among plant TPSs including many bacterial-like TPSs containing just the α domain (Jia *et al.* 2016, Blerot *et al.* 2018, Kumar *et al.* 2018, Wei *et al.* 2019)

To date, the Terpene_Synth_C Pfam domain (PF03936) only contains 106 sequences of mono-, sesqui- and diterpene bacterial TPSs from 83 bacterial species. This is much fewer than the 262 sequences predicted by Yamada *et al.* (2015) and it also contains none of the class I thermostable TPSs mentioned above. This might be attributed to the low sequence identity of bacterial class I TPSs or the frequency in which profile-HMMs are performed by the Pfam database. This suggests that independent profile-HMMs, combined with the increased number of sequenced bacterial genomes will continue to find new TPS sequences that are not currently in the PF03936 family. The primary aim of this work was to find more thermostable TPSs for use in the *Parageobacillus* terpene platform to make it more industrially valuable. The regular running of new HMMs with a focus on genome sequences from thermophiles should identify potential new thermophilic class I TPSs which can be interrogated to expand the pool of terpenes that are produced naturally at high temperatures. To this effect, new independent profile-HMMs were run throughout this project, by using a diverse range of sequences of both classes of bacterial TPS as inputs. These included monodomain (Class I, α domain) and didomain (Class II, γβ) TPSs. Selected class I sequences from thermophiles would then be characterised for both thermostability and their terpene product.

3.2 Results and Discussion

3.2.1 HMMER searches for new thermophilic terpene synthases

Bacterial class I TPSs have up to four well characterised motifs within a sequence of ~350 residues. In their order in the primary sequence, these are defined as the Asp-rich motif, DDX(X)(D/E), the pyrophosphate sensor, R, the NSE triad, (N/D)(D/E)XX(S/T)XX(N/K/R)(D/E) and the RY motif, RY. The Asp-rich, NSE triad and the RY motifs coordinate the three Mg^{2+} ions and the diphosphate of the substrate to enable correct binding (Aaron *et al.* 2010, Baer, Rabe, Fischer, *et al.* 2014). The pyrophosphate sensor detects binding of the diphosphate moiety of the substrate to allow the reaction to be initiated by a small conformational change (Baer, Rabe, Citron, *et al.* 2014, Baer, Rabe, Fischer, *et al.* 2014). Class II diterpene synthases have a different domain structure and also have three conserved motifs. The Asp-rich motif (DXDD) is found in the β-domain where the neighbouring aspartates in this motif are vital for the protonation of the carbon-carbon double bond to initiate cyclisation (Prisic *et al.* 2007). There are also two truncated QW motifs (QXXDG(G/S)WG), one in each domain on either side of the Asp-rich motif (Feil *et al.* 1996, Peters *et al.* 2001, Cao *et al.* 2010). Triterpene cyclases (squalene-hopene and oxidosqualene cyclases) also contain the γβ didomain structure of class II diterpene synthases but in addition to the Asp-rich motif in the β-domain as well as between 5-8 full length QW motifs ((R/K)(A/G)XX(X)(F/W/Y)LXXXQXXXGXW) instead of the truncated version. Both the truncated and full length QW motifs sit at the base of the α-helices and interact with each other to stabilize the protein during its exergonic cyclisation reaction (Wendt *et al.* 1997, Siedenburg and Jendrossek 2011). These were not included in the search parameters but were predicted to be present in the search result in large numbers as TriTPSs are present in most bacteria and share sequence identity with class II bacterial diterpene synthase. Both of these enzyme classes have low sequence similarities and identities so sequences were collected for both classes for use in separate HMMs.

3.2.1.1 Bacterial Class I TPS HMM Search

To search for more thermostable bacterial class I TPSs, 15 protein sequences were collected from the NCBI database (Table 3.1). To maximise the chances of returning

sequences with higher sequence variation from the class I HMM, sequences were collected that had low sequence homologies and less identity in the motifs. The overall sequence identities varied from 10.2-28.8%. The alignments of the motifs of the collected sequences are shown in Figure 3.1. In several cases, the third aspartate/glutamate residue of the Asp-rich motif was separated by one or two residues from the other aspartate residues. The only sequences that showed an unusual Asp-rich motif were odyverdiene A synthase, which had a shorter four residue motif with alternating glutamate and aspartate residues, and intermedeol synthase which lacked the third aspartate residue altogether. The pyrophosphate sensor and the RY motif were conserved in all sequences except the (-)-pimara-9(11),15-diene and intermedeol synthases. In (-)-pimara-9(11),15-diene, the pyrophosphate sensor was replaced with an alanine while the tyrosine of the RY motif was an aspartate, while in intermedeol synthase phenylalanine replaced the tyrosine of the RY motif. In the key positions of the NSE triad, the residues were mainly conserved except for the asparagine position in α-amorphene synthase which was replaced by a proline. These different TPS also take on substrates of different sizes i.e. GPP, FPP and GGPP to produce a wide range of terpene structures spanning mono-, sesqui- and diterpenes (Figure 3.2).

Table 3.1: Bacterial class I α-domain TPSs used for the monodomain HMM. The terpene products are shown in Figure 3.2. [1] The first three residues of α-amorphene synthase residues were not included in the alignment. [2] As no gene names were identified, the Uniprot Accession Numbers were provided instead.

Organism	Terpene Product	Gene Name	References
Streptomyces clavuligerus	δ-Cadinene	*screlav_p0328*	(Hu *et al.* 2011)
Streptomyces sp. ND90	Odyverdiene A	*nd90_0354*	(Yamada *et al.* 2015, 2016)
Streptomyces viridochromogenes	α-Amorphene	*ssqg_00965*[1]	(Rabe and Dickschat 2013)
Pseudanabaena limnetica str. Castaic Lake	2-Methylisoborneol	*e9nh28*[2]	(Giglio *et al.* 2011)
Streptomyces lactacystinaeus	(-)-Isohirsut-4-ene	*slt18_1880*	(Yamada *et al.* 2015, 2016)
Streptomyces clavuligerus ATCC 27064	Intermedeol	*sclav_p0635*	(Yamada *et al.* 2015)
Streptomyces clavuligerus ATCC 27064	African-2-ene	*sclav _p0985*	(Yamada *et al.* 2015)
Streptomyces clavuligerus ATCC 27064	Linalool/Nerolidol	*sclav _p1185*	(Nakano, Kim, *et al.* 2011a)
Roseiflexus sp. RS-1	τ-Muurolol	*roseRS_3509*	(Yamada *et al.* 2015, Styles *et al.* 2017)
Kitasatospora griseola	Terpentetriene	*cyc2_Kitgr*	(Dairi *et al.* 2001)
Streptomyces sp. KO-3988	(-)-Pimara-9(11),15-diene	*q5ksn4*[2]	(Kawasaki *et al.* 2006)
Kitasatospora setae KM-6054	Corvol Ether B	*kse_12950*	(Rabe *et al.* 2015)
Saccharopolyspora spinosa NRRL 18395	Hedycaryol	*a8926_4163*	(Dickschat *et al.* 2014)
Streptosporangium roseum	4-*epi*-Cubebol	*sros_6866*	(Dickschat *et al.* 2014)
Streptomyces sp. OX=1931	Cembrene C	*dtcycA*	(Meguro *et al.* 2013)

	Motifs			
Terpene Synthase	**Asp-rich**	**PP Sensor**	**NSE Triad**	**RY**
2-Methylisoborneol	DDYYADD	R	NDLLSVAKD	RY
4-epi-cubebol	DDAF-CE	R	NDLISYAKE	RY
African-1-ene	DEQF-DD	R	NDIVSLPKE	RY
A-amorphene	DDRA-ED	R	PDLFSAVKE	RY
Cembrene_C	DDLR-FE	R	NELYSAGKE	RY
Corvol_Ether_B	DDLF-VD	R	NDVQSLKME	RY
Δ-cadinene	DDRI-DE	R	NDLMTVDKE	RY
Hedycaryol	DDSL-DR	R	NDLVSARNE	RY
Intermedeol	DDLA-LP	R	NDIVSYERE	RF
Linalool	DDQF-DA	R	NELHSFEKD	RY
Odyverdiene_A	EDED-CG	R	NDTHSLERE	RY
Pimaradiene	DDHV-EQ	A	NDLATFERE	RD
T-muurolol	DDQC-DE	R	NDMLSYPKE	RY
Terpentetriene	DDRW-DY	R	NDYYSWGRE	RY
(-)-Isohirsut-4-ene	DDAL-DE	R	NDIVSLEKD	RY
	::		: : :	*

Figure 3.1: Aligned Asp-rich (DDXX(X)D), pyrophosphate (PP) sensor (R), NSE triad (N(D/E)XX(S/T)XX(D/E)) and the RY motifs of the selected class I bacterial TPSs for the HMM. The full alignment is shown in Appendix Figure 3. The percentage similarity and identity of each of these genes is shown in Appendix Figure 2.

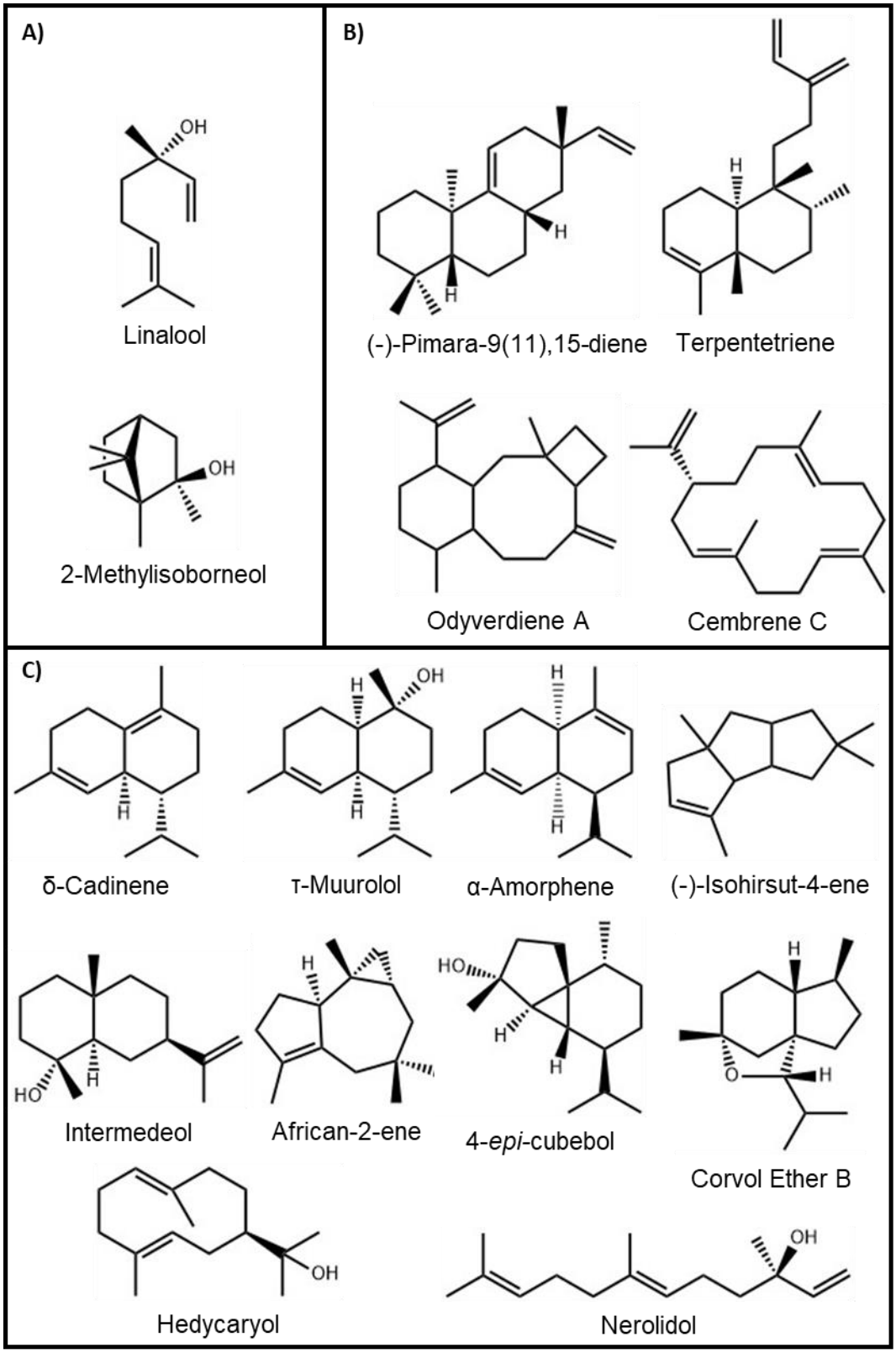


Figure 3.2: A) mono-, B) di- and C) sesquiterpene structures of the products from the TPSs used in the class I HMM from GPP, GGPP and FPP, respectively. For most of the TPSs, only a single terpene is shown but both linalool and nerolidol are displayed for the linalool/nerolidol synthase.

The sequence alignment of the class I bacterial TPSs was input into the profile-HMMER web server with the E-value cut-off score set as 10^{-5} as E-values higher than this returned increased numbers of false positives. The profile-HMMER was run against the UniProtKB sequence database with the search being restricted to only bacterial sequences collecting 4517 sequences. Of these sequences, 431 had to be removed as they corresponded to fragments of a TPS sequence. Any sequences shorter than 250 aa were also removed as they are unlikely to be TPSs at this length. The list of sequences identified by the Class I HMM is on the attached supplementary media under the filename “Class I TPS HMM Results.” Sequences with higher E-values were checked for the appropriate motifs, at which point several sequences were removed for not having either the Asp-rich motif or the NSE triad. Of the remaining 4086 returned sequences, with E-values between $9.2e^{-6}$ and $6.2e^{-151}$, most were not labelled with any pfam domains, and only ~240 sequences were labelled with the Terpene_synth_C (PF03936) pfam domain that is indicative of the α-domain of class I TPSs. Several sequences corresponded to TPSs that had already been characterised for their terpene product. However due to their low sequence identity and similarity, most terpene products cannot be deduced from their TPS primary sequence or structure alone, with some notable exceptions such as geosmin synthases (Yamada *et al.* 2015). The majority of sequences were ~350 residues in size as expected; however, there were a number of sequences that were much larger. This included several multi-domain proteins containing extra Pfam domains that are associated with terpene synthesis including Cytochrome P450 (PF00067) or Polyprenyl Synthetase (PF00348) domains. Some of the longer sequences (<900 sequences), with lengths up to approximately 800aa, were likely to be geosmin/germacredienol synthases that contain two α-domains, with all of the motifs, as opposed to just one. Many sequences with very similar E-values were predicted to be duplicates. In many cases, this was due to the sequences arising from highly related species. This might have led to slight over-representation of certain phyla or even genera like *Streptomyces* as each strain naturally contains three or more TPS genes (Jiang *et al.* 2007). However, it would not be possible to account for this over-representation without product analysis of every TPS from each individual species. Regardless, Actinobacteria dominated the resulting sequences accounting for >75%, of which two thirds were from *Streptomyces*. The remaining ~1000 sequences came, in descending order, from Proteobacteria, Chloroflexi, Cyanobacteria, Firmicutes, Acidobacteria, Deinococcus-Thermus and Nitrospirae. The individual TPSs from Deinococcus-Thermus and Nitrospirae are the first of their kind. Of the 4085 sequences, only 22 sequences derived from thermophilic bacteria (Table 3.2).

Table 3.2: Returned thermophilic bacterial TPS sequences from the Class I HMM Search. The sequence from *K. papyrolyticum* was not identified using the HMM but instead through a BLAST search with RoseRS_3509. Any growth temperatures that were not described were left as N/A (Not available). [1] *R. radiotolerans* was not tested for growth above 55 °C although it grew very well at 55 °C. [2] *T. staphylospora* range was based on *T. grisealba*. [3] This temperature was the highest temperature these strains grew at with no minimum or optimum temperatures tested. Growth temperature of *F. thermalis* CCMEE 5330 was predicted from the other isolates as no temperature was recorded. [4] This temperature for *F. thermalis* was that of the location of the bacterial isolates. [5] For the *K. papyrolyticum* sequence, the NCBI accession number was used as no gene name was available.

Organism	Optimum Temperature (°C)	Temp Range (°C)	Gene Name	Length (aa)	E-values	Phylum	Reference
Amycolatopsis eurytherma	N/A	25-55	*edd35_0439*	728	$2.1e^{-131}$	Actinobacteria	(Kim *et al.* 2002)
			edd35_0260	348	$6.5e^{-67}$		
			edd35_4817	325	$1.8e^{-34}$		
Roseiflexus sp. RS-1	55-60	45-60	*roseRS_3509*	326	$7.0e^{-103}$	Chloroflexi	(van Der Meer *et al.* 2010)
			roseRS_2648	323	$3.7e^{-52}$		
Roseiflexus castenholzii DSM 13941	50	45-60	*rcas_0622*	326	$1.9e^{-88}$	Chloroflexi	(van Der Meer *et al.* 2010)
			rcas_2716	324	$9.9e^{-54}$		
Rubrobacter xylanophilus DSM 9941	60	40-70	*rxyl_0493*	324	$1.6e^{-77}$	Actinobacteria	(Carreto *et al.* 1996)
Rubrobacter radiotolerans	46-48	37-55[1]	*rradSPS_2540*	315	$1.2e^{-74}$	Actinobacteria	(Yoshinaka *et al.* 1973)
			rradSPS_0143	348	$5.2e^{-42}$		

Thermostaphylospora chromogena	50	40-60[2]	*samn04489764_5014*	335	$7.3e^{-72}$	Actinobacteria	(McCarthy and Cross 1984, Wu *et al.* 2018)
Fischerella thermalis CCMEE 5205	N/A	57[3]	*cen40_22155*	328	$8.5e^{-65}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5318	N/A	57[3]	*cen46_16690*	328	$5.3e^{-65}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis	51-54[4]	N/A	*fjsc11draft_4294*	328	$1.4e^{-64}$	Cyanobacteria	(Brown *et al.* 2007)
Fischerella thermalis CCMEE 5201	N/A	55[3]	*cen39_03285*	328	$3.6e^{-64}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5268	N/A	57[3]	*cen50_12075*	328	$1.5e^{-64}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5273	N/A	56[3]	*cen49_08730*	328	$8.7e^{-64}$	Cyanobacteria	(Miller *et al.* 2007)

Fischerella thermalis CCMEE 5330	N/A	55[3]	*cen41_20415*	328	$7.3e^{-63}$	Cyanobacteria	(Miller *et al.* 2007)
Thermomonospora curvata DSM 43183	50	40-65	*tcur_3107*	347	$8.2e^{-42}$	Actinobacteria	(Chertkov *et al.* 2011)
			tcur_3806	316	$5.0e^{-10}$		
Thermobispora bispora DSM 43833	N/A	50-65	*tbis_3257*	292	$6.3e^{-07}$	Actinobacteria	(Liolios *et al.* 2010)
Kallotenue papyrolyticum JKG1^{T}	55	40-65	*wp_029215342*	326	-	Chloroflexi	(Cole *et al.* 2013)

The majority of the retrieved sequences from thermophiles also came from the phylum Actinobacteria. The only sequences that were from different phyla were those from *Roseiflexus* species, which are Chloroflexi, and *Fischerella* species which are Cyanobacteria. Of the 22 sequences identified, RoseRS_3509, Rcas_0662 and Rxyl_0493 have been previously identified as τ-muurolol and cembrene C synthases, respectively (Smanski *et al.* 2012, Yamada *et al.* 2015) but the structural stability of Rxyl_0493 against temperature has not been previously studied (Styles *et al.* 2017). Tcur_3107 was also previously identified but displayed no activity *in vivo* in the engineered *S. avermitilis* system (Yamada *et al.* 2015). The rest of the sequences have not been identified or characterised previously. The additional sequence in Table 3.2 from *Kallotenue papyrolyticum* JKG1^{T}, referred to as JKG1 for the rest of this study, was identified from a BLAST search with RoseRS_3509 as the query sequence and is also from the phylum Chloroflexi. However, compared to RoseRS_3509, it has only 35.3% identity and 54.2% similarity. The sequence identity of these enzymes is fairly low meaning the terpene product cannot be predicted. However, EDD35_0439 was predicted to be a geosmin/germacradienol synthase based on its size, having the conserved motifs in both the N- and C-terminal domains as well as showing 54.3% sequence identity in a 753-residue alignment with the geosmin synthase (SCO6073) from *S. coelicolor* (Jiang *et al.* 2007). The products of the rest of the sequences could not be predicted, although it is likely that all of the *Fischerella* sequences produce the same compound as each sequence was the same length and the E-values do not vary greatly. Expression and characterisation studies were required to determine the terpenes these enzymes produce.

3.2.1.2 Bacterial Class II TPS HMM Search

Alongside the bacterial class I (α) HMM, a class II (γβ) diterpene synthase HMM was set up to compare sequences returned of the alternative type of TPS in bacteria as performed with the class I HMM. Many class II diterpene synthases are present in plants, but fewer bacterial examples exist (Smanski *et al.* 2012). This meant that only five class II diterpene synthase sequences were collected with five unique, albeit similarly structured terpene products (Table 3.3 and Figure 3.3). Although the sequence identities of these enzymes are slightly higher than the class I HMM, they are still <30%. It is worth noting that this HMM focussed on enzymes that use GGPP. There are also class II diterpene synthases that use epoxy-GGPP as a substrate in

which the Ap-rich motif is replaced with a (E/D)SA(E/N) motif. These enzymes make intermediates on the way to forming phenalinolactone A (*Streptomyces* sp. Tü6071) and brasilicardin A (*Nocardia terpenica*) and they are believed to be related to oxidosqualene cyclases used for sterol production in eukaryotes (Dürr *et al.* 2006, Hayashi *et al.* 2008). These were not included in this HMM because there are relatively few examples.

Table 3.3: Bacterial diterpene synthases used for the γβ didomain HMM.

Organism	Terpene Product	Gene Name	Reference
Mycobacterium tuberculosis	Tuberculosinyl diphosphate	*rv3377c*	(Nakano and Hoshino 2009)
Kitasatospora griseola	Terpentedienyl diphosphate	*cyc1_kitgr*	(Dairi *et al.* 2001)
Bradyrhizobium diazoefficiens USDA 110	Ent-copalyl diphosphate	*blr2149*	(Morrone *et al.* 2009)
Herpetosiphon auranticus	(+)-kolavenyl diphosphate	*haur_2145*	(Nakano *et al.* 2015)
Salinispora arenicola CNS-205	Copalyl diphosphate	*sare_1288*	(Xu *et al.* 2014)

Terpene Synthase	QW Motif 1	Asp-rich	QW Motif 2
Blr2149	QQADGGWG	DADD	QRDDGGWG
Cyc1_KitGr	QRPDGLWG	DGDD	QTDSGGWG
Haur_2145	QQSDGGWG	DGDD	QHPDGGWG
Rv3377c	QLPDGSWG	DCDT	QRPDGSWG
Sare_1288	QRPDGGWG	DADT	QRANGSWG
	* .** **	* *	* .*.**

Figure 3.3: Sequence alignment of the motifs of the diterpene synthases used for the class II HMM. The full alignment is shown in the Appendix Figure 4. The percentage similarity and identity of each of these genes is shown in Appendix Figure 2.

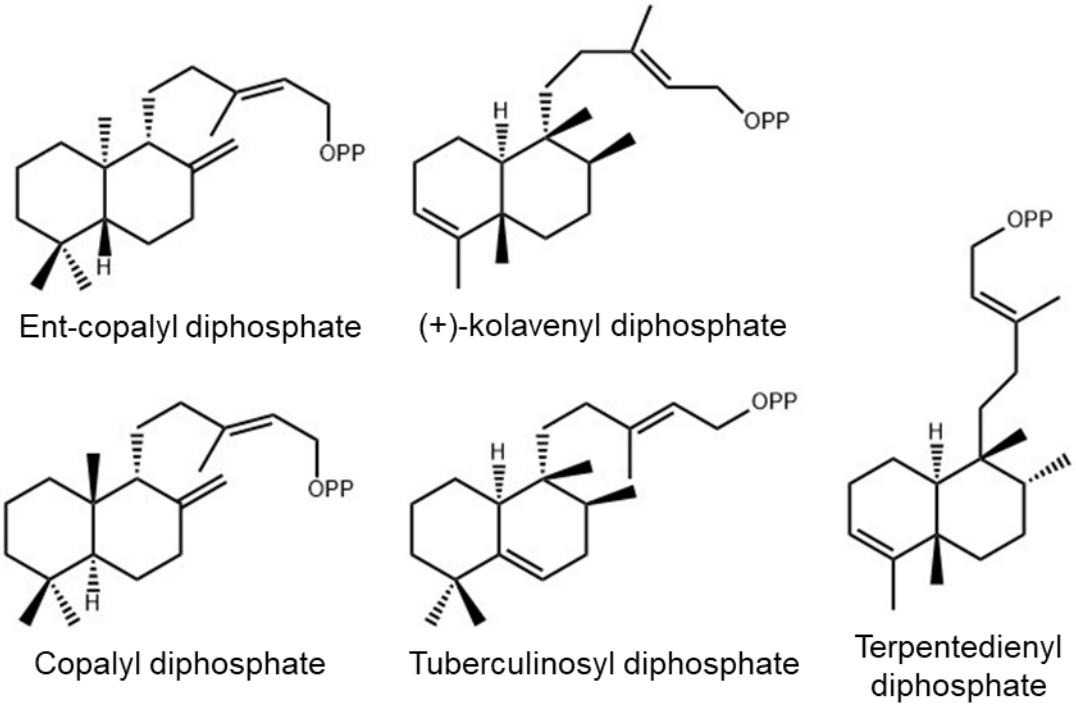


Figure 3.4: Diterpene diphosphates produced from GGPP by the TPSs in Table 3.3 used for the class II bacterial HMM.

As with the class I HMM, the sequence alignment was inserted into the profile-HMMER server to search the UniProtKB database with cut-off E-values for hit scores of 10^{-5} and the search taxonomy restricted to bacteria only. This search returned a total of 5363 sequences after partial TPS sequences, labelled as fragments, had been removed. The full list of sequences is on the attached supplementary media under the filename “Class II TPS HMM Results”. E-values ranged from 1.3 e^{-223} to 9.6 e^{-06}. Sequences with the lowest E-values between 1.3 e^{-223} to 3.1 e^{-36}, and therefore the highest similarity to the search alignment, were predicted to be class II diterpene synthases. From the search this accounted for 489 after sequences shorter than 425 aa had been removed as they were too short to represent this enzyme group. Most sequences were between 425-626 residues in length and contained the truncated QW motifs and the Asp-rich domain. Only five phyla were represented in this result with Proteobacteria and Actinobacteria representing more than 90% with the remaining sequences coming from Chloroflexi, Cyanobacteria and Firmicutes in descending order, respectively. The majority of these enzymes were predicted to contain the pfam domains of either both the Prenyltrans (PF00432) and SQHop_cyclase_C (PF13243), just SQHop_cyclase_C or no pfam domains at all.

The Prenyltrans pfam domain is a short region that contains the conserved truncated QW motifs.

The remaining 4858 sequences had E-values between 1.9 e^{-31} and 8.9 e^{-07} with most of the sequences predicted to be squalene cyclases (SQCs) or similar with the majority of these containing the pfam domains, SQHop_cyclase_N (PF13249) and SQHop_cyclase_C indicative of SQCs. Sequences that were shorter than 520 aa, or described as prenyltransferases were removed. The sequences predicted to be prenyltransferases either lacked enough QW motifs to be SQCs, had the Asp-rich motif in the incorrect position or both. Any sequences with unique descriptions were checked for the appropriate motifs before keeping or discarding them. This left 4323 sequences thought to be SQCs or related enzymes. The SQC sequences come from all of the phyla predicted to contain diterpene synthases, in addition to Acidobacteria, Armatimonadetes, Bacteroidetes, Elusimicrobia, Fibrobacteres, Gemmatimonadetes, Lentisphaerae, Nitrospinae, Nitrospirae, Planctomyces, Spirochaetes and Verrucomicrobia with several unlabelled sequences. Actinobacteria, Firmicutes and Proteobacteria accounted for over 75% of the sequences. The most significant SQC sequence (1.9 e^{-31}) from the HMM search was from *Alicyclobacillus acidocaldarius* and contains all eight QW motifs with the central Asp-rich motif; it has been characterised previously (Dang and Prestwich 2000). As the E-values got higher, the sequences had between 5-8 QW motifs and total lengths varying between 523-800 aa, with several longer sequences containing extra domains. Although function could not be predicted for most of the SQC sequences, specific types of SQC were predicted which included oxidosqualene cyclases (oSQCs) and tetraprenyl-β-curcumene cyclases. When compared to the *A. acidocaldarius* SQC, oSQCs have a DCTAE motif instead of the Asp-rich motif, where the final aspartate residue from the Asp-rich motif becomes the first residue of the DCTAE motif. Tetraprenyl-β-curcumene cyclases only have six truncated QW motifs and the Asp-rich motif. They are similar in length to diterpene synthases. Tetraprenyl-β-curcumene cyclases and oSQCs use tetraprenyl-β-curcumene and (2,3)-oxidosqualene, respectively, as substrates as opposed to squalene (Frickey and Kannenberg 2009, Sato *et al.* 2011). Although many oSQCs have been identified in bacteria, more eukaryotic oSQCs and their sterols have been characterised (Wei *et al.* 2016, Banta *et al.* 2017). Similarly to the class I HMM, species that have a lot of strains sequenced appear more regularly, especially as the products of SQCs are vital in primary metabolism and therefore exist in all related strains. As well as 596 SQC sequences being predicted from 525 unique *Streptomyces* strains alone, 754 SQCs are predicted from 451 unique *Bacillus*

strains. This may inflate the number of sequences from specific phyla, in this case Actinobacteria and Firmicutes, but also outlines the importance of SQCs in these phyla due to being present in most sequenced strains.

In terms of class II TPSs, the HMM shows that there are many more SQCs than diterpene synthases and that they are more diversely spread across the bacterial domain. Although the study of these enzymes is limited, the products of SQCs, called hopanoids, have multiple vital functions, which is why they occur more often in bacteria than diterpene synthases. Their main function is to stabilize the membrane and maintain its fluidity similar to that of sterols in eukaryotes (Poralla *et al.* 1984, Nalin *et al.* 2000). In comparison, functions of the resulting diterpenes are specific to certain species rather than having a common purpose. Several examples are isotuberculisinol, gibberellins, platencin and platensimycin. Isotuberculisinol is a virulence factor in *M. tuberculosis* that promotes survival in macrophages (Mann *et al.* 2009, Nakano *et al.* 2011). Gibberellins are produced by plant-associated bacteria, mainly rhizobia, and facilitate plant-bacterial interactions as well as improving plant growth (Bottini *et al.* 2004, Nett *et al.* 2017). Over a third of the diterpene synthase genes predicted from the HMM come from rhizobia. These are likely involved in gibberellin production from *ent*-copalyl diphosphate. Platencin and platensimycin are antibiotics from *Streptomyces platensis* (Wang *et al.* 2006, 2007). These functions would mean that fewer diterpene synthases should be expected by this search as they are uncommon. The numbers of SQCs compared to diterpene synthases is also reflected in the number sequences from mild and true thermophiles returned, with 85 SQC sequences coming from 10 bacterial phyla and eight diterpene synthase sequences coming from two phyla but mainly from Actinobacteria (Tables 3.4 and 3.5). The only characterised thermophilic class II TPS is that of the SQC from *A. acidocaldarius* (Dang and Prestwich 2000). Of the thermophilic class II diterpene synthases, most were previously identified by Yamada *et al.* (2012) except *Saccharomonospora xinjiangensis* XJ-54, *Litorilinea aerophila* and those from *Amycolatopsis* species. It is likely that the product of the diterpene synthases from *Saccharomonospora viridis* DSM 43017 and *Saccharomonospora xinjiangensis* XJ-54 is the same. The *Amycolatopsis* diterpene synthases also probably produce the same diterpene diphosphate as one another. This is due to the highly similar E-values but characterisation is still required to confirm this.

Table 3.4: Predicted diterpene synthase sequences from mild thermophiles and thermophiles from the class II TPS HMM. Any temperatures that could not be identified were labelled as not available (N/A). [1] *S. xinjiangensis* was not tested at a temperature higher than 50 °C and therefore might have a wider growth temperature range.

Organism	Optimum Temperature (°C)	Temp Range (°C)	Accession Number	Length (aa)	E-values	Phylum	Reference
Thermobispora bispora (strain ATCC 19993 / DSM 43833 / CBS 139.67 / JCM 10125 / NBRC 14880 / R51)	N/A	50-65	Tbis_3256	531	$3.0e^{-120}$	Actinobacteria	(Liolios *et al.* 2010)
Amycolatopsis methanolica 239	N/A	10-50	AMETH_2896	509	$2.5e^{-119}$	Actinobacteria	(de Boer *et al.* 1990)
Amycolatopsis eurytherma	N/A	25-55	EDD35_0136	509	$4.9e^{-117}$	Actinobacteria	(Kim *et al.* 2002)
Saccharomonospora viridis DSM 43017	55	37-60	SAMN02982918_3177	534	$5.2e^{-115}$	Actinobacteria	(Pati *et al.* 2009)
			Svir_08360	530	$1.1e^{-114}$	Actinobacteria	
Saccharomonospora xinjiangensis XJ-54	45-50	28-50[1]	SacxiDRAFT_1154	547	$3.2e^{-113}$	Actinobacteria	(Jin *et al.* 1998)
Thermomonospora curvata DSM 43183	50	40-65	Tcur_3808	545	$7.0e^{-110}$	Actinobacteria	(Chertkov *et al.* 2011)
Litorilinea aerophila	55	40-65	RY27_18830	519	$1.8e^{-108}$	Chloroflexi	(Kale 2013)

Table 3.5: Mildly thermophilic and thermophile SQCs predicted from the class II HMM. Any growth temperatures that could not be identified were labelled as not available (N/A). [1] The temperature range for Acetothermia bacterium 64_32 was that of where it was isolated. [2] As no values are quoted for *A. acidocaldarius* LAA1, values from *A. acidocaldarius* subsp. *acidocaldarius* DSM 446 were used. [3] *S. xinjiangensis* was not tested at a temperature higher than 50 °C and therefore might have a wider growth temperature growth range. [4] This temperature was the highest temperature these strains grew at with no minimum or optimum temperatures tested. *F. thermalis* CCMEE 5330 was predicted from the other isolates as no temperature predicted. [5] This temperature was that of the location of the bacterial isolates. [6] *Thermoflavimicrobium dichotomicum* was not tested at a temperature higher than 55 °C and therefore might have a wider growth temperature range. [7] Temperature at which these strains were grown at.

Organism	Optimum Temperature (ºC)	Temp Range (ºC)	Accession Number	Length (aa)	E-values	Phylum	Reference
Alicyclobacillus acidocaldarius subsp. *acidocaldarius* DSM 446	60-65	45-70	P33247	631	$1.9e^{-31}$	Firmicutes	(Mavromatis *et al.* 2010)
Alicyclobacillus acidocaldarius subsp. *acidocaldarius* Tc-4-1	70	N/A	F8IIV2	631	$2.4e^{-30}$	Firmicutes	(Chen *et al.* 2011)
Acetothermia bacterium 64_32	N/A	47-70[1]	A0A3B9QQ07	643	$3.3e^{-30}$	Acetothermia	(Hu *et al.* 2016)
			A0A351JLX5	645	$1.1e^{-27}$		
Alicyclobacillus acidocaldarius LAA1[2]	60-65	45-70	B7DNH1	631	$8.8e^{-30}$	Firmicutes	(Mavromatis *et al.* 2010, Chen *et al.* 2011)

Alicyclobacillus vulcanalis	55	35-65	A0A1N7JW90	631	$1.6e^{-29}$	Firmicutes	(Simbahan *et al.* 2004)
Sulfobacillus thermosulfidooxidans	50-55	20-60	A0A2T2WP40	559	$7.3e^{-29}$	Firmicutes	(Bogdanova *et al.* 2006)
Saccharomonospora xinjiangensis XJ-54	45-50	28-50[3]	I0V8F5	568	$8.8e^{-27}$	Actinobacteria	(Jin *et al.* 1998)
			I0V7I9	638	$1.4e^{-23}$		
Alicyclobacillus tengchongensis	45	30-50	A0A0V0PRS5	633	$5.7e^{-26}$	Firmicutes	(Kim *et al.* 2014)
			A0A0V0PNG4	623	$3.0e^{-18}$		
Sphaerobacter thermophilus DSM 20745	55	N/A	D1C5C0	617	$6.4e^{-26}$	Chloroflexi	(Pati *et al.* 2010)
Thermodesulforhabdus norvegica	60	44-74	A0A1I4R6V2	673	$2.2e^{-25}$	Proteobacteria	(Beeder *et al.* 1995)
Acidothermus cellulolyticus 11B	55	37-70	A0LVK9	633	$3.8e^{-25}$	Actinobacteria	(Mohagheghi *et al.* 1986)
Alicyclobacillus hesperidum	50-53	35-60	A0A1H2R2P0	633	$7.3e^{-25}$	Firmicutes	(Albuquerque *et al.* 2000)
			A0A1H2Y0M6	623	$2.4e^{-18}$		
Alicyclobacillus hesperidum URH17-3-68	50-55	N/A	J9HMR6	622	$1.8e^{-24}$	Firmicutes	(Wang *et al.* 2012)
			J9HEP4	625	$5.6e^{-19}$		
Saccharomonospora viridis DSM 43017	55	37-60	A0A1I5JV38	640	$2.7e^{-24}$	Actinobacteria	(Pati *et al.* 2009)
			C7MXN7	640	$2.7e^{-24}$		

Amphiplicatus metriothermophilus	48-50	37-65	A0A239PQ81	680	$7.9e^{-24}$	Proteobacteria	(Zhen-Li *et al.* 2014)
Alicyclobacillus acidoterrestris ATCC 49025	42-53	25-60	T0DEU9	633	$1.1e^{-22}$	Firmicutes	(Deinhard *et al.* 1987, Shemesh *et al.* 2013)
			T0CIK5	604	$4.8e^{-10}$		
			T0DCL2	621	$1.1e^{-08}$		
Amycolatopsis methanolica 239	N/A	10-50	A0A076MTU2	633	$1.9e^{-21}$	Actinobacteria	(de Boer *et al.* 1990)
Thermasporomyces composti	50-55	35-62	A0A3D9V7T2	672	$2.7e^{-21}$	Actinobacteria	(Yabe *et al.* 2011)
Amycolatopsis eurytherma	N/A	25-55	A0A3N2H5Z2	633	$8.5e^{-21}$	Actinobacteria	(Kim *et al.* 2002)
Thermogutta terrifontis	60	30-60	A0A286RF18	734	$3.9e^{-20}$	Planctomycetes	(Slobodkina *et al.* 2015)
			A0A286RG14	653	$5.4e^{-15}$		
Chloracidobacterium thermophilum (strain B)	51	44-58	G2LJM4	659	$1.2e^{-19}$	Acidobacteria	(Tank and Bryant 2015)
Thermosynechococcus sp. NK55a	N/A	52-60	V5V678	642	$2.9e^{-19}$	Cyanobacteria	(Stolyar *et al.* 2014)
Thermosynechococcus elongatus BP-1	55	N/A	Q8DGK8	642	$9.9e^{-19}$	Cyanobacteria	(Nakamura *et al.* 2002)
Methylacidiphilum fumariolicum SolV	55	50-60	I0K028	689	$1.1e^{-18}$	Verrucomicrobia	(Mohammadi *et al.* 2016)

Methylacidiphilum infernorum V4	60	N/A	B3DYT5	691	$2.0e^{-16}$	Verrucomicrobia	(Hou *et al.* 2008)
Chthonomonas calidirosea T49	68	50-73	S0EVH8	692	$4.0e^{-16}$	Armatimonadetes	(Lee *et al.* 2011)
Fischerella thermalis CCMEE 5205	N/A	57[4]	A0A2N6MTD1	636	$5.6e^{-15}$	Cyanobacteria	(Miller *et al.* 2007)
Synechococcus lividus PCC 6715	63-67	54-72	A0A2D2Q4H4	640	$5.9e^{-15}$	Cyanobacteria	(Meeks and Castenholz 1971)
Fischerella thermalis	51-54[5]	N/A	G6FPH9	636	$1.8e^{-14}$	Cyanobacteria	(Brown *et al.* 2007)
Fischerella thermalis CCMEE 5273	N/A	56[4]	A0A2N6KQB0	636	$1.8e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5330	N/A	55[4]	A0A2N6M9X8	636	$2.5e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5268	N/A	57[4]	A0A2N6KFV0	636	$2.7e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5318	N/A	57[4]	A0A2N6LCN3	636	$2.8e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)
Fischerella thermalis CCMEE 5319	N/A	57[4]	A0A2N6LQA3	636	$2.8e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)

Fischerella thermalis CCMEE 5201	N/A	55[4]	A0A2N6N491	636	$3.0e^{-14}$	Cyanobacteria	(Miller *et al.* 2007)
Bacillus licheniformis	51	N/A	Q62TY2	629	$2.0e^{-13}$	Firmicutes	(Warth 1978)
			A0A3P2APH1	629	$4.3e^{-13}$		
			A0A3S5DIS9	629	$4.2e^{-13}$		
			A0A1Y0YPJ1	629	$1.2e^{-12}$		
			A0A1Q9FEN9	632	$1.6e^{-09}$		
Parageobacillus thermoglucosidasius	61-63	37-68	A0A1B7KUL3	621	$4.5e^{-13}$	Firmicutes	(Suzuki *et al.* 1983)
Aeribacillus pallidus	55	30-70	A0A165XT78	620	$5.3e^{-13}$	Firmicutes	(Scholz *et al.* 1987, Yasawong *et al.* 2011)
Thermoactinomyces sp. AS95	55	40-65	A0A151Z081	630	$6.7e^{-13}$	Firmicutes	(Bezuidt *et al.* 2016)
Thermoactinomyces vulgaris	55	N/A	A0A0N0HCD4	624	$1.4e^{-11}$	Firmicutes	(Reijula 1993)
Geobacillus sp. 44B	55[5]	N/A	A0A1V9APN9	621	$1.4e^{-11}$	Firmicutes	(Singh, Carlson, *et al.* 2017)
Geobacillus sp. 8	60	30-65	A0A164B284	620	$1.8e^{-11}$	Firmicutes	(Shintani *et al.* 2014)

Laceyella sediminis	55	28-65	A0A2T0YC03	631	$1.8e^{-11}$	Firmicutes	(Chen, Lin, *et al.* 2012)
Bacilllus methanolicus MGA3	50-53	37-60	I3E8W8	627	$2.7e^{-11}$	Firmicutes	(Schendel *et al.* 1990)
Bacilllus methanolicus PB1	55	35-60	I3DVK1	587	$2.9e^{-11}$	Firmicutes	(Arfman *et al.* 1992)
Bacillus sonorensis L12	N/A	15-55	M5PHV6	629	$3.4e^{-11}$	Firmicutes	(Palmisano *et al.* 2001, Adimpong *et al.* 2013)
Parageobacillus caldoxylosilyticus	65	38-75	A0A150LMW1	612	$7.0e^{-11}$	Firmicutes	(Ahmad *et al.* 2000)
Parageobacillus caldoxylosilyticus NBRC 107762	N/A	N/A	A0A023DK02	628	$1.6e^{-10}$	Firmicutes	N/A
Anoxybacillus sp. B7M1	60	40-65	A0A160FD04	624	$3.0e^{-10}$	Firmicutes	(Filippidou *et al.* 2016)
Paenibacillus thermophilus	42-45	37-60	A0A388MND1	634	$3.9e^{-10}$	Firmicutes	(Zhou, Gao, Wei, *et al.* 2012)
Geobacillus sp. MAS1	N/A	N/A	V6VLG1	618	$4.4e^{-10}$	Firmicutes	N/A
Bacillus tequilensis	N/A	25-50	A0A2X0X2Z1	633	$4.8e^{-10}$	Firmicutes	(Gatson *et al.* 2006)

Geobacillus sp. Y412MC52	65	55-75	A0A0E0TCC4	618	$7.1e^{-10}$	Firmicutes	(Brumm *et al.* 2015)
Geobacillus thermocatenulatus	60	37-80	A0A226QDQ1	618	$1.5e^{-08}$	Firmicutes	(Dinsdale *et al.* 2011)
Geobacillus thermoleovorans	60	37-70	A0A2Z3NCS4	618 617	$7.8e^{-10}$ $1.4e^{-08}$	Firmicutes	(Dinsdale *et al.* 2011)
Thermoflavimicrobium dichotomicum	N/A	30-55[6]	A0A1I3LGK3	620	$1.2e^{-09}$	Firmicutes	(Yoon *et al.* 2005)
Geobacillus genomosp. 3 (sp. JF8)	60	30-75	S5Z030	617	$1.5e^{-09}$	Firmicutes	(Shintani *et al.* 2014)
Geobacillus zalihae	65	50-70	A0A1V9BZM7 A0A1V9CM17	618 618	$1.8e^{-09}$ $1.1e^{-09}$	Firmicutes	(Rahman *et al.* 2007)
Geobacillus sp. GHH01	N/A	N/A	L7ZZ94	617	$2.1e^{-09}$	Firmicutes	N/A
Geobacillus sp. LEMMY01	55[7]	N/A	A0A1V4PB21	618	$3.0e^{-09}$	Firmicutes	(Pinheiro Alves de Souza *et al.* 2017)
Geobacillus kaustophilus GBlys	55[7]	N/A	U2YCZ7	618	$3.0e^{-09}$	Firmicutes	(Doi *et al.* 2013)*
Geobacillus sp. 46C-IIa	55[5]		A0A1V9B3B4	617	$3.4e^{-09}$	Firmicutes	(Singh, Carlson, *et al.* 2017)

Geobacillus sp. PA-3	N/A	N/A	A0A0Q0YYM9	617	$3.6e^{-09}$	Firmicutes	N/A
Geobacillus thermodenitrificans	50	45-70	A0A1W6VSM4	617	$4.2e^{-09}$	Firmicutes	(Manachini *et al.* 2000, Nazina *et al.* 2001)
Geobacillus thermodenitrificans NG80-2	65	45-73	A4IPI8	617	$4.2e^{-09}$	Firmicutes	(Feng *et al.* 2007)
Lihuaxuella thermophila	50	28-65	A0A1H8JBI1	626	$5.9e^{-09}$	Firmicutes	(Yu *et al.* 2012)
Geobacillus sp. T6	N/A	N/A	A0A0J0V4P7	617	$1.2e^{-08}$	Firmicutes	N/A
Geobacillus subterraneus	55-60	45-65	A0A143ML35	618	$1.5e^{-08}$	Firmicutes	(Poli *et al.* 2013)
Bacillus coagulans	N/A	25-60	A0A150KG99	648	$3.9e^{-08}$	Firmicutes	(Nakamura *et al.* 1988)
Melghirimyces profundicolus	50-55	37-65	A0A2T6C8R1	631	$6.9e^{-08}$	Firmicutes	(Li *et al.* 2013)
Melghirimyces thermohalophilus	55	43-60	A0A1G6NX08	637	$1.9e^{-07}$	Firmicutes	(Nariman Addou *et al.* 2013)

3.2.1.3 Comparison of HMM Result

From the HMMs, it was shown that class I TPSs exist in the same phyla as class II diterpene synthases with several individual exceptions for the class I TPSs in Acidobacteria, Bacteroidetes, Chlamydiae, Deinococcus-Thermus and Nitrospirae. For class II diterpene synthases, this is unsurprising as the cyclised diphosphate products of these enzymes tend to be cyclised further by a class I (Wang *et al.* 2006, 2007, Mann *et al.* 2009, Nakano *et al.* 2011). Class I enzymes can also cyclise GPP, FPP and GGPP to produce a wider variety of terpene structures than the class II diterpene synthases. There were over 8-fold more class I TPSs compared to class II diterpene synthases. When class I TPSs were compared to SQCs, the majority of the species that contained a class I TPS also contained a SQC but not *vice versa*. SQCs were also spread across more phyla than the class I TPSs. Even though similar numbers of sequences were identified as class I TPSs and SQCs, class I TPS sequences were predominantly found in Actinobacteria (>75% of sequences obtained), with multiple TPSs being found in each Actinobacteria species. This meant that there was a two-fold difference in the number of individual species that contained SQCs compared to class I TPSs with 2885 species versus 1371 species, respectively. The phyla Armatimonadetes, Elusimicrobia, Fibrobacteres, Gemmatimonadetes, Lentisphaerae, Nitrospinae, Planctomycetes, Spirochaetes and Verrucomicrobia all contained SQCs but no class I TPSs. This suggests that many related bacterial species will contain a single SQC but generally not more than one. With the exception of Actinobacteria, having a class I TPS is probably not characteristic of a phylum, but each strain may have multiple class I TPSs.

The formation of many diterpenes requires pairs of class I and II TPS to form a single terpene product. One example mentioned previously was the cyclisation of GGPP to *ent*-copalyl diphosphate by the class II diterpene synthase, PtmT2, followed by a second cyclisation to platensimycin by the class I TPS, PtmT3 (Wang *et al.* 2006, Smanski *et al.* 2012). No class II diterpene synthase has been identified without a class I partner previously. From the HMMs, of the thermophilic sequences, only *T. bispora* DSM 43833 and *T. curvata* DSM 43183 had possible class I and II TPS gene pairs encoded by *tbis_3257/tbis_3256* and *tcur_3806/tcur_3808*, respectively, which probably act together to produce a single product. The remaining thermophilic class II diterpene synthases did not have a class I TPS partner but did have a neighbouring cytochrome P450 enzyme in the genome. P450 enzymes normally perform further modifications after cyclisation. The lack of a class I TPS in the genome brings into

question how the diphosphate is removed from the product of the class II diterpene synthase and whether further cyclisations are required. Currently, no class II diterpene synthase has been characterised without a class I TPS partner. The genome sequence for *L. aerophila* is incomplete so may contain a class I TPS even though one has not been found. Also, *A. eurytherma* contains three class I TPSs, none of which neighbour the class II diterpene synthase in the genome. As one of these is probably a geosmin synthase, one of the other two could possibly interact with the product of the class II diterpene synthase.

Of the three enzymes types identified from these HMMs, only SQCs and class I TPSs have enzymes fully characterised from thermophiles (Siedenburg and Jendrossek 2011, Styles *et al.* 2017). However, this HMM showed that more SQC examples exist in thermophiles. The importance of SQC products (Hopanes) in thermophiles has been shown, however, the function of the only class I TPS product from a thermophile, τ-muurolol, is unknown (Llopiz *et al.* 1996, Garcia Costas *et al.* 2012, Caron *et al.* 2014). The small number of thermophilic class I TPSs might be the result of their low sequence identity. However, it is more probable that it reflects the increased volatility of the monoterpenes and sesquiterpenes at high temperatures. This would limit the number of functions that volatile terpenes can usefully perform, implying that there are likely to be fewer class I TPSs present in these thermophiles. It should be noted that, no TPS sequences were predicted to be present in hyperthermophiles; class I TPSs were highly unlikely to be present in these organisms for the reasons outlined previously. Furthermore hyperthermophiles do not use SQC products to stabilize their membranes explaining the lack of these sequences in the HMM results (Siliakus *et al.* 2017)

3.2.2 Expression of putative thermophilic terpene synthases

All of the class I thermostable TPSs model as single α-domains when put through both SWISS-MODEL and I-TASSER. However, before choosing which sequences to characterise from Table 3.2, they were aligned and checked to confirm the presence of the conserved motifs of a class I TPS (Figure 3.5). For the remainder of this study, FJSC11DRAFT_4294 and SAMN04489764_5014 are referred to as FJSC11 and Tchrom, respectively.

	Motifs			
Terpene Synthase	**Asp-rich**	**PP Sensor**	**NSE Triad**	**RY**
CEN39_03285	DDY-YDT	R	NDIYSIPKE	RY
CEN40_22155	DDY-YDT	R	NDIYSIPKE	RY
CEN41_20415	DDY-YDT	R	NDVYSIPKE	RY
CEN46_16690	DDY-YDT	R	NDIYSIPKE	RY
CEN49_08730	DDY-YDT	R	NDIYSIPKE	RY
CEN50_12075	DDY-YDT	R	NDIYSIPKE	RY
EDD35_0260	DDARCDS	R	NDLHSYGKE	RY
EDD35_0439	DDH-FLE	R	NDLFSYQRE	RY
EDD35_4817	DDF-LDE	R	NDLFSYQRE	RY
FJSC11	DDY-YDT	R	NDIYSIPKE	RY
JKG1	DDL-CSA	R	NDIISLAKE	RY
Rcas_0622	DDQ-CDE	R	NDVLSYPKE	RY
Rcas_2716	DDL-TDT	R	NDLWSYPKE	RY
RoseRS_2648	DDL-TDT	R	NDLWSYPKE	RY
RoseRS_3509	DDQ-CDE	R	NDMLSYPKE	RY
RradSPS_0143	DDR-WDA	R	NDLVSLNKE	RY
RradSPS_2540	DDL-CDE	R	NDIISLTKE	RY
Rxyl_0493	DDL-ADA	R	NDIISLAKE	RY
Tchrom	DDQ-FEL	K	NDVVGLRRE	RF
Tbis_3257	DWQ-AER	A	NDLATSGRE	--
Tcur_3107	DDH-FDD	R	NDVVSYPKE	RY
Tcur_3806	DWI-MDH	A	NDLATWNRD	EG
	*		**: ::	

Figure 3.5: Multiple sequence alignment of the conserved motifs of the putative class I TPSs in Table 3.2 predicted by the HMM. The N-terminal domain of the predicted geosmin synthase, EDD35_0439, was used for the sequence alignment as either the N-terminal or C-terminal could be used.

Most sequences contained all four motifs; however, several key differences were observed. JKG1 lacked the third aspartate residue from the Asp-rich motif, which although rare, has been identified in other TPSs; selina-11-ene-4α-ol and dauca-8,11-diene synthase have the Asp-rich motifs of DDLAL and DDYFA, respectively (Yamada *et al.* 2015, Dickschat 2016). On the other hand, Tbis_3257 and Tcur_3806 replace the second aspartate residue with a tryptophan which has not been seen in a TPS to date. The dependence of catalysis on the aspartate residues varies from synthase to synthase, however, the second aspartate has consistently been shown to be vital for catalysis with its mutation causing a drastic reduction in activity, which makes these enzymes interesting (Seemann *et al.* 2002, Felicetti and Cane 2004, Aaron *et al.* 2010, Baer, Rabe, Fischer, *et al.* 2014, Grundy *et al.* 2016). If these enzymes are active, the tryptophan residue may still enable coordination of the Mg^{2+} ions and the diphosphate through the use of its conjugated ring system. No significant changes were seen in the pyrophosphate sensor as the arginine was switched to a lysine in Tchrom and Tbis_3257 and Tcur_3806 had an alanine similar to that seen in the search sequence, (-)-pimara-9(11),15-diene synthase. The only difference in the NSE triad is that Tchrom had a glycine in the central position instead of a

serine/threonine. This position is believed to coordinate one of the three Mg^{2+} ions however several examples of this change exist in plant TPSs, including kaurene synthase from *Arabidopsis thaliana* and *Oryza sativa* (Xu *et al.* 2006, Zhou and Peters 2009). Mutations of the serine/threonine residues to glycine in this position have been shown to decrease catalytic activity but maintain function, however this enzyme may have adapted to this change (Zhou and Peters 2009, Aaron *et al.* 2010, Grundy *et al.* 2016). Tchrom also replaces the tyrosine of the RY motif with a phenylalanine which was present in the search sequence, intermedeol synthase. Switching the tyrosine residue with a phenylalanine has been shown to maintain wild-type activity (Baer, Rabe, Fischer, *et al.* 2014); however the reverse change has never been tested. The RY motif aids coordination of both the diphosphate moiety and a Mg^{2+} ion. Tbis_3257 and Tcur_3806 lack this motif entirely, and therefore cannot provide its coordination. For Tbis_3257, this is a consequence of having a much shorter sequence than the other bacterial TPSs. Both of these potential TPSs are worth investigating further to try and understand how these enzymes completely exclude the active site from the solvent.

Of the sequences in Table 3.2, the genes from *Thermomonospora curvata (tcur_3107)*, *Fischerella thermalis* (*fjsc11*), *Thermostaphylospora chromogena* (*tchrom*), *Thermobispora bispora* (*tbis_3257*), *Rubrobacter xylanophilus* (*rxyl_0493*) and *Kallotenue papyrolyticum* (*jkg1*) were selected to be expressed for characterisation. Rxyl_0493 was purified to test for thermostability as it has previously had its activity characterised (Citron *et al.* 2012, Yamada *et al.* 2015) while the other sequences were tested for both the terpene product and thermostability of the enzyme. Tcur_3107 was purified for *in vitro* assays as Yamada *et al.* (2015) could not see activity in the *S. avermitilis* system. Prior to gene synthesis, the closest characterised homologues to Tcur_3107, FJSC11, Tchrom, Tbis_3257 and JKG1 were identified (Table 3.6). All of the potential TPSs other than Tbis_3257 and Tchrom were most similar to, and had high coverage with sesquiterpene synthases. Tbis_3257 and Tchrom showed the highest similarity to the diterpene synthase, ent-pimara-9(11),15-diene synthase, and the monoterpene synthase, 1,8-cineole synthase, respectively. While most had sequence identities below 35%, FJSC11 had over 50% identity to germacrene A from *Nostoc punctiforme*. Although high, this level of similarity does not necessarily suggest that these two enzymes perform the same function as difference in individual residues can affect product formation. While, it is more likely that the closest relative naturally utilises the same substrate as the putative TPS analysis is still required to determine the terpene product.

Table 3.6: Characterised TPSs with the highest sequence identities and similarities to the selected potential TPSs according to a Uniprot BLAST search. The coverage percentage was the length of the matched sequence over the length of the query sequence. Accession numbers: (+)-caryolan-1-ol synthase - B1W019, germacrene A synthase - B2J4A4, 1,8-cineole synthase - B5GMG2 and ent-pimara-9(11),15-diene synthase - Q5KSN4.

Terpene Synthase	Most Similar to	Organism	Identity (%)	Similarity (%)	Coverage (%)
Tcur_3107	*(+)-Caryolan-1-ol synthase*	*Streptomyces griseus subsp. griseus*	30.3	40.3	96.5
FJSC11	Germacrene A synthase	*Nostoc punctiforme* strain ATCC 29133	50.8	71.0	98.4
Tchrom	1,8-cineole synthase	*Streptomyces clavuligerus* DSM 738	32.9	47.1	98.5
Tbis_3257	Ent-pimara-9(11),15-diene synthase	*Streptomyces* sp. strain KO-3988	25.9	40.4	99.0
JKG1	Germacrene A synthase	*Nostoc punctiforme* strain ATCC 29133	34.8	54.2	99.1

The gene sequences of the six prospective TPSs were codon optimized for expression in *E. coli,* cloned into pET28a between the NheI and EcoRI sites and transformed into *E. coli* Bio*Blue*. For expression using the T7 promoter, the pET28a:TPS constructs were transformed into *E. coli* BL21(DE3). Expression conditions were optimised for each TPS and then expression was scaled up and each TPS purified using a N-terminal His tag by affinity chromatography with a cobalt resin. For the TPS, *rxyl_0493* was produced in 2L of culture, *tchrom* and *jkg1* in 2.5L culture and *tbis_3257*, *tcur_3107* and *fjsc11* in 3L of culture. After purification, the proteins were concentrated to ~5 mg/ml. SDS-PAGE gels are shown in Appendix Figure 11 of each TPS purified by affinity chromatography. The TPS, *jkg1,* had to be expressed with an extra 24 residues on the C-terminus by removing the stop codon in the plasmid otherwise the enzyme aggregated when concentrated. Similarly, only small quantities of Rxyl_0493 as it also aggregated under concentration. No FJSC11 could be retrieved after being concentrated as it had aggregated out of solution so the second wash was buffer exchanged as it contained relatively pure FJSC11. Carlos Amaya (Former MSc student, Leak laboratory) cloned, expressed and purified JKG1 and Tchrom under my close supervision.

To confirm that these proteins form the predominantly α-helical structure characteristic of TPSs, wavelength scans using CD were performed. The scans

(Figure 3.6) of the five TPSs, showed the double negative minima at 208 nm and 222 nm expected of a primarily α-helical structure suggesting proper folding (Greenfield 2007). Unfortunately, a complete wavelength scan could not be completed for Rxyl_0493 as increased noise produced, due to the lower dilution in in buffer, lowered the reliability of the data, however it had already been confirmed as having TPS-activity so likely had an α-helical bundle structure. The next step was to establish whether any of the TPSs were thermostable. Each TPS was subjected to a CD thermal melt to determine the melting point (T_m) of their secondary structures (Figure 3.7 and Table 3.7).

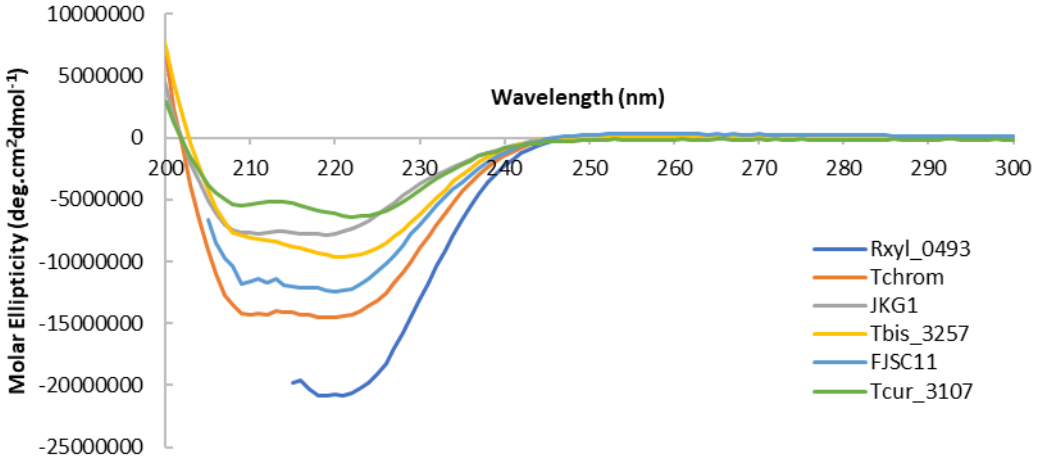


Figure 3.6: Wavelength scans of Tcur_3107 Tbis_3257, JKG1, FJSC11 and Tchrom following the molar ellipticity of the secondary structures at 20 °C using CD. Tchrom, Tbis_3257 and Tcur_3107 were measured between 200-300 nm and FJSC11 between 205-300 nm. Rxyl_0493 was not able to have a full wavelength scan performed and was only between 215-300 nm. Due to the lower dilutions of FJSC11 and Rxyl_0493 in CD buffer, the wavelength scans could not be run lower than 205 nm and 215 nm, respectively to avoid the noise increasing and therefore lowering the reliability of the data.

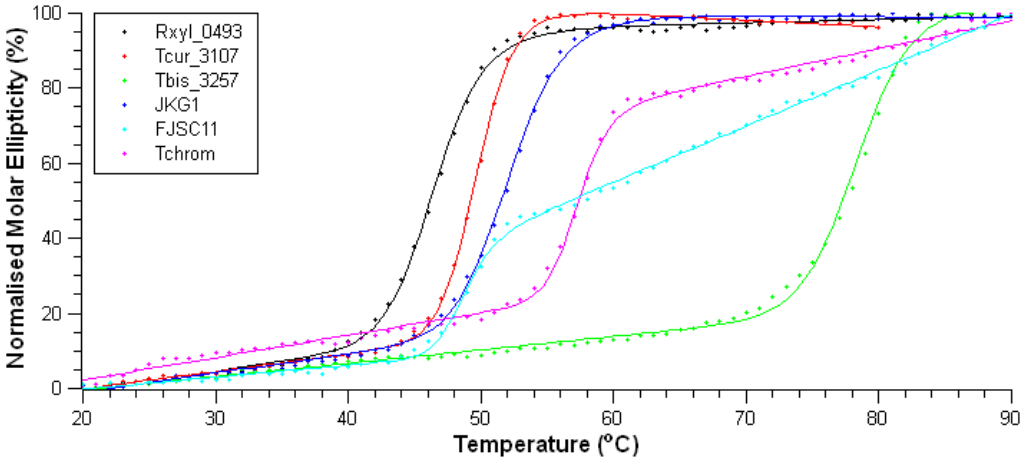


Figure 3.7: CD thermal melts between 20-90 °C of Rxyl_0493, Tcur_3107, Tbis_3257, JKG1, FJSC11 and Tchrom following the molar ellipticity of the secondary structures at 222 nm. The melt for Tcur_3107 was only performed up to 80 °C rather than 90 °C. The R^2 values were 0.998 for Rxyl_0493 and Tbis_3257 and 0.999 for Tchrom, JKG1, FJSC11 and Tcur_3107 and represent the fit of the data to the curve generated using Equation 4 (Chapter 2.6.5).

Table 3.7: T_m and van't Hoff enthalpy of unfolding (ΔH) values of Rxyl_0493, Tcur_3107, Tbis_3257, JKG1, FJSC11 and Tchrom calculated from the thermal melts. The error is calculated as the deviation of the data from the curve made by Equation 4 (Chapter 2.6.5).

Terpene Synthase	T_m (°C)	ΔH (kJ/mol)
Rxyl_0493	46.5 ± 0.04	52.3 ± 1.0
Tcur_3107	49.7 ± 0.03	74.3 ± 1.1
Tbis_3257	78.5 ± 0.09	58.4 ± 2.3
JKG1	52.2 ± 0.05	51.3 ± 0.9
FJSC11	48.5 ± 0.12	77.0 ± 6.2
Tchrom	57.0 ± 0.11	70.6 ± 4.4

The T_m values suggested that all of the enzymes were at least mildly thermostable with Rxyl_0493 having the lowest value. Most of the T_m values sat in the temperature range for growth of their associated organisms (Table 3.2), however the T_m of Tbis_3257 was 78.5 °C, well above the growth temperature of *T. bispora*, making it highly thermostable.

Next, to determine what type of TPS each enzyme was and its associated product, each purified enzyme was incubated with GPP or FPP separately for 1 hour at 50 °C. Tcur_3107 was incubated at 37 °C for 1 hour. Although, some unfolding may have occurred during these assays, the identification of the product was the priority, therefore any potential unfolding was not considered. Assays were performed in

triplicate and analysed by GC-MS analysis on the hexane extracts. The structure of each terpene peak on the GC-MS spectra was predicted by the NIST database and confirmed using analytical standards where possible (Figures 3.8-3.10). As Rxyl_0493 had already been characterised as a cembrene C synthase, it was not examined again. All of the monoterpenes produced could be confirmed using authentic standards (Appendix Figure 5) but only (+)-sativene could be confirmed for the sesquiterpenes peaks as no authentic standard was available for the remaining sesquiterpenes (Appendix Figure 6).

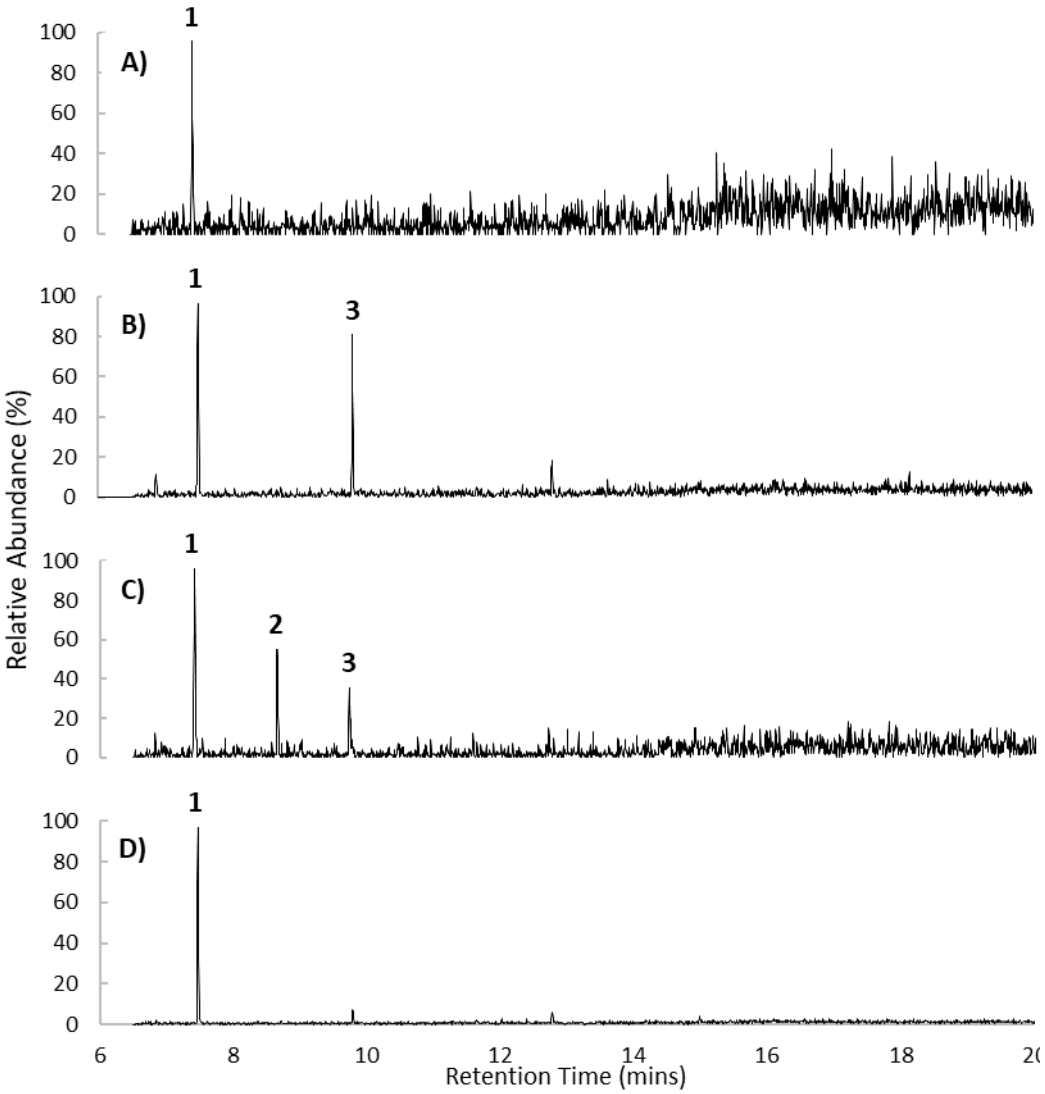


Figure 3.8: GC-MS chromatograms of the monoterpenes produced by A) Tbis_3257, B) JKG1, C) FJSC11 and D) Tchrom. The three monoterpene products from the GPP assays were determined using the extracted ion content at 93 m/z and their retention times were 1) linalool at 7.419 mins, 2) α-terpineol at 8.656 mins and 3) geraniol at 9.744 mins.

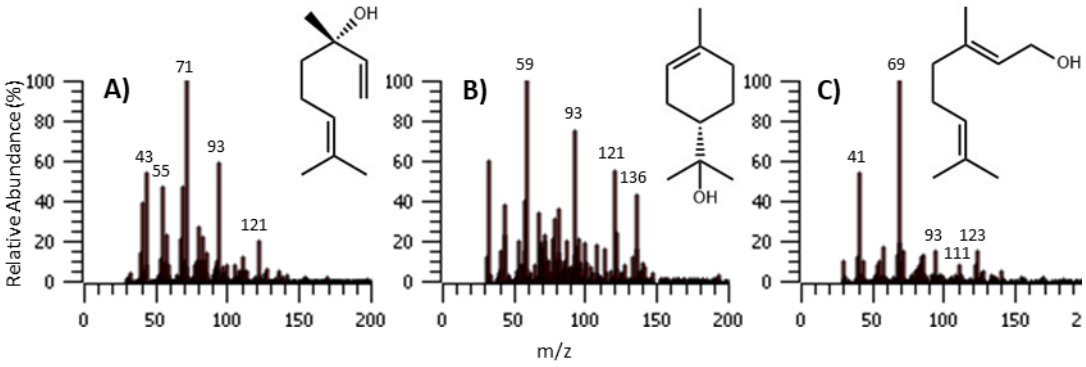


Figure 3.9: Examples GC-MS mass spectrum of the monoterpene products from Tchrom, JSC11 and JKG1. The monoterpenes were A) linalool from Tchrom, B) α-terpineol from JSC11 and C) geraniol from JKG1 corresponding to 1, 2 and 3 from Figure 3.8, respectively.

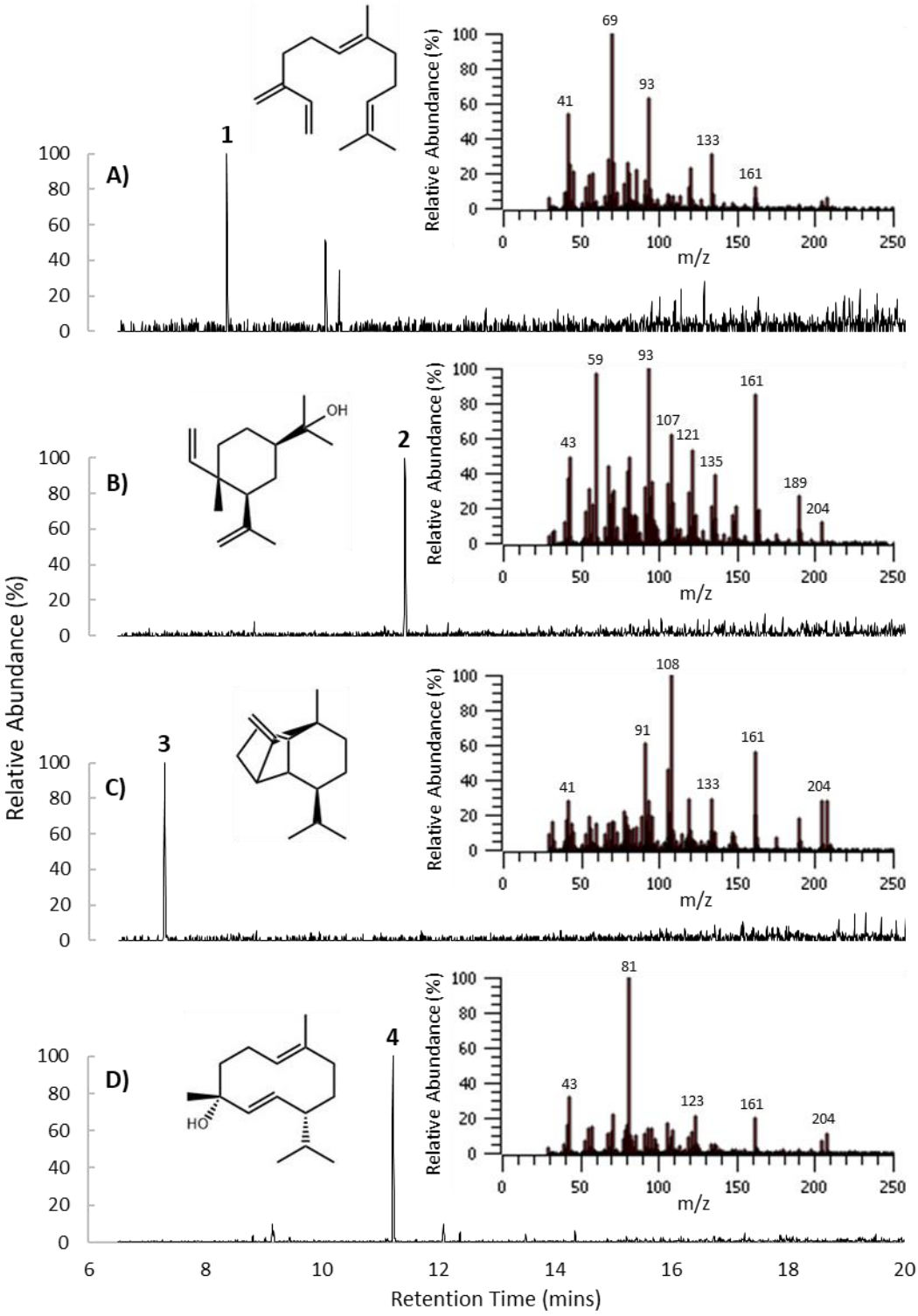


Figure 3.10: GC-MS chromatograms, mass spectrums and structures of the sesquiterpenes produced A) Tbis_3257, B) JKG1, C) FJSC11 and D) Tchrom. The four different products were determined using the XIC at 204 m/z of FPP assays. The retention times of each sesquiterpene was 1) β-farnesene at 8.376 mins, 2) β-elemol at 11.427 mins, 3) (+)-sativene at 7.294 mins and 4) germacrene D-4-ol at 11.236 mins.

From the GC-MS spectra, Tbis_3257 produced linalool from GPP and β-farnesene from FPP, however, the spectra appeared to show more of the former being produced under the same conditions. Both were confirmed using authentic standards. Linalool is formed by the addition of water to either the geranyl cation or linalyl cation at the C3 position to produce the tertiary alcohol (Schnee *et al.* 2002, Srividya *et al.* 2016). On the other hand, β-farnesene is likely to be produced by deprotonation of the farnesyl cation at position C3 (Crock *et al.* 1997). As both of these products are acyclic, these terminations occur on diphosphate removal from the substrate before any movement of the carbocation preventing any cyclisation. However, rather than being labelled a rare bacterial monoterpene synthase like the linalool/nerolidol synthase from *S. clavuligerus* (Nakano, Kim, *et al.* 2011a), Tbis_3257 is likely to be a diterpene synthase. In the *T. bispora* genome, Tbis_3257 is transcribed with Tbis_3256 and a cytochrome P450, Tbis_3258. Tbis_3256 was predicted by the class II HMM in this study and Yamada et al (2012) as a diterpene synthase. This transcriptional coupling suggests that these two enzymes work one after the other to cyclise GGPP as no class II mono- or sesquiterpene synthases have ever been identified. The cytochrome P450 probably oxidises the product of the two cyclisation steps. This would be similar to the partially characterised gene cluster that produces terpentecin in *Kitasatospora griseola* (Dairi *et al.* 2001). It contains both class II (Cyc1) and I (Cyc2) TPSs for converting GGPP into terpentedienyl diphosphate and then terpentetriene. From here, two cytochrome P450s probably perform the final steps in the formation of terpentecin. Both Cyc1 and Cyc2 were used as search parameters in the HMMs above. However, whereas Cyc2 contains all four of the conserved motifs, Tbis_3257 lacks both the pyrophosphate sensor and the RY motif making them significantly different. Similarly to Tbis_3257, Cyc2 produced small amounts of β-farnesene when assayed with FPP but unlike Cyc2, Tbis_3257 was able to convert GPP to linalool (Hamano *et al.* 2002). Without performing assays containing Tbis_3256 and Tbis_3257 both individually and together, it is not possible to ascertain the terpene structures produced by these reactions and provide further evidence that these are diterpene synthases. Kinetics with GPP and FPP were not performed on this enzyme as it was predicted from the time course and quantity of terpenes produced in the previous GC-MS assays that this enzyme would be very slow.

Through *in vivo* studies by Yamada *et al.* (2012) and *in vitro* assays in this study, the mildly thermostable Tcur_3107 still showed no activity despite having the four motifs characteristic of a class I TPS and modelling like one (No data shown). This suggests that it might take on another substrate, however, it is unlikely to be GGPP as a

diterpene should have been observed by Yamada *et al.* (2012) when using the *S. avermitilis* SUKA22 strain. As discussed, when considering the class II HMM, it is possible that Tcur_3107 binds other substrates such as *ent*-copalyl diphosphate or terpentedienyl diphosphate. However, for this to be the case, it would typically form an operon with a class II diterpene cyclase to produce a substrate for it as seen with the terpentecin cluster from *K. griseola* and the kolavelool cluster from *H. aurantiacus* (Dairi *et al.* 2001, Nakano *et al.* 2015). There is a class II diterpene synthase in *T. curvata*, Tcur_3808, predicted by this study. However, this is ~0.76 Mbp away from Tcur_3107 and more likely interacts with the other predicted class I TPS from *T. curvata*, Tcur_3806 and a cytochrome P450, Tcur_3807, which it neighbours (Table 3.2). However, unlike the *T. bispora* cluster, only Tcur_3806 and Tcur_3807 are predicted to be transcriptionally coupled. As with the TPSs from *T. bispora*, individually assaying and co-assaying Tcur_3806 and Tcur_3808 may be necessary to observe any TPS activity from *T. curvata*. It is unlikely that Tcur_3107 forms part of this pathway. It could cyclise a modified precursor but for now its activity remains uncharacterised.

The NIST database predicted that JKG1 cyclised GPP to linalool and geraniol and FPP to β-elemol. The identities of monoterpenes were confirmed using authentic standards. The mechanism for linalool formation has been described previously for Tbis_3257. Similar to the linalool mechanism, geraniol is likely produced from either linalyl cation or the geranyl cation where the hydroxyl is added to C1 producing the acyclic primary alcohol (Schnee *et al.* 2002). As both of these linear monoterpenes are produced, this suggests that on carbocation formation the addition of water happens non-specifically on either C1 or C3. Over the time course of an hour, much more β-elemol was produced than the monoterpenes suggesting that this is a sesquiterpene synthase. However, β-elemol is not likely to be the final product of this enzyme. Studies have shown that β-elemol arises from thermal rearrangement of hedycaryol caused by the high temperatures in the GC-MS (Figure 3.11) (Jones and Sutherland 1968, Hattan *et al.* 2016). This suggests that JKG1 is a hedycaryol synthase. Hedycaryol has been previously identified in the Actinobacteria, *Kitasatosporae setae* and *Saccharopolyspora spinosa*, where it is likely formed through conversion of FPP to NPP, 1,10-cyclisation of NPP and hydroxyl addition to C11 (Baer, Rabe, Citron, *et al.* 2014, Dickschat *et al.* 2014). This would be the first Chloroflexi example of a hedycaryol synthase. As it was not thermostable enough to be used in the *Geobacillus* system, it was not characterised kinetically.

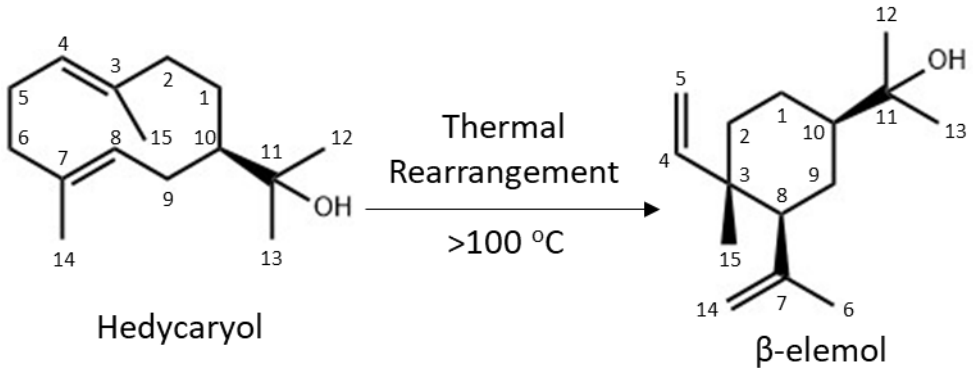


Figure 3.11: Thermal rearrangement of hedycaryol to β-elemol at temperatures above 100 °C.

FJSC11 was fairly promiscuous when incubated with GPP, producing three monoterpenes. Similarly, to JKG1 it produced the linear linalool and geraniol, but it was the only tested TPS to produce a cyclic monoterpene, α-terpineol. To form α-terpineol, it is likely that the linalyl cation is converted into the terpinyl cation by a 1,6-ring closure followed by electrophilic addition of water to the cation at C7 (Martin and Bohlmann 2004). This suggests that this enzyme enabled the polyprenyl chain to be folded in the active site of FJSC11 to enable cyclisation instead of immediate termination to form linalool or geraniol like the other TPSs in this study. With FPP, it was predicted to only produce (+)-sativene, which was confirmed with an authentic standard. The mechanism for (+)-sativene cyclisation has only been predicted computationally. This mechanism is believed to begin by conversion of FPP to NPP, 1,10-cyclisation of NPP followed by a hydride transfer, two further cyclisations between carbons 1,6 and 3,7 and an alkyl transfer before termination by deprotonation (Lodewyk *et al.* 2008). In nature, sativene has only been identified from plants and fungi. However, rather than being the dominant product of the reaction, it forms as a side product for enzymes such as γ-humulene synthase from *Abies grandis* and δ-cadinene from *Coprinus cinereus* (Steele *et al.* 1998, Agger *et al.* 2009). To date, it has not been isolated from bacteria. More investigation is required including the use of NMR to confirm the structure and GC-MS of deuterated reaction product to determine the mechanism as well as kinetics. Kinetics were not performed as this TPS could not be used in the *Parageobacillus* terpene system due to its low T_m value.

Tchrom only produced a single product from GPP which was linalool, similarly to Tbis_3257. With FPP, it produced mainly germacrene D-4-ol with small amounts of

τ-muurolol/α-cadinol making it a germacrene D-4-ol synthase under *in vitro* conditions. Germacrene D-4-ol has been predicted to form through a 1,10-cyclisation of NPP or FPP with a 1,3-hydride shift followed by the addition of water at C3 to produce the monocyclic alcohol. However, due to the production of tau-muurolol/alpha-cadinol, Tchrom probably cyclises via the NPP route. FPP can be converted to NPP by the sesquiterpene synthase (Jones *et al.* 2008, Grundy *et al.* 2016). As no analytical standard was available, NMR analysis could be used to determine structure of the terpene product and GC-MS using deuterated FPP to determine the mechanism of cyclisation. This would help to confirm that Tchrom is a germacrene D-4-ol synthase. As this enzyme was a fairly thermostable TPS, kinetics with both GPP and FPP were performed at 50 °C with the malachite green assay (Table 3.8). As the assays were performed at 50 °C, the unfolding of Tchrom may have been convoluted but as unfolding is difficult to follow in these assays, it was not measured.

Table 3.8: Kinetics of Tchrom with GPP and FPP using the malachite green assay. Concentrations used were 0-136.9 µM for GPP and 0-92.4 µM for FPP. Michaelis-Menten kinetic curves are shown in Appendix Figure 16. Errors shown are standard errors.

Substrate	K_M (µM)	k_{cat} (s^{-1})	k_{cat}/K_M ($x10^{-3}$ $s^{-1}mM^{-1}$)
GPP	9.7 ± 2.2	0.009 ± 0.001	0.9 ± 0.2
FPP	9.8 ± 2.3	0.069 ± 0.005	7.1 ± 1.7

Although the K_M is similar for GPP and FPP, the k_{cat} is around 8-fold higher for FPP resulting in a higher catalytic efficiency. This suggests that this enzyme is a germacrene D-4-ol synthase rather than a linalool synthase. The catalytic efficiency of the only other characterised bacterial linalool synthase from *Streptomyces citricolor* was 6.5 $mM^{-1}s^{-1}$ which is very similar to this enzyme. Compared to plant linalool synthases, the catalytic efficiency of Tchrom is similar or higher (Crowell *et al.* 2002, Landmann *et al.* 2007, Nagegowda *et al.* 2008). As a germacrene D-4-ol synthase, the catalytic efficiency of Tchrom is lower than that of the only other kinetically characterised bacterial germacrene D-4-ol synthase from *Streptomyces citricolor* which is 1.72 μM^{-1} s^{-1} (Nakano *et al.* 2011), however the kinetic parameters of Tchrom are comparable to other bacterial TPSs (Agger *et al.* 2008, Hu *et al.* 2011, Styles *et al.* 2017).

3.3 Conclusions

Three prior iterations of HMMs searching for new TPSs increased the number of putative class I TPSs from 41 to 262 as well as finding 33 new class II diterpene synthases (Komatsu *et al.* 2008, Yamada *et al.* 2012, 2015). In this chapter, this technique was performed to sample the increasing number of sequenced bacterial genomes for class I and class II TPSs focussing on those from thermophilic organisms, especially class I TPSs. This allowed for the diversity across the bacterial domain of the TPS classes to be directly compared for the first time using newly made sequence alignments as search parameters. As expected, this analysis showed that class II SQCs are present in more bacteria than class I TPSs or class II diterpene synthases as well as being spread across a wider range of bacterial phyla, a pattern which was also seen in the of thermophilic TPS sequences identified. As SQCs were not part of the initial search alignment, they may not have all been identified by this search.

Of the predicted thermophilic class I and class II TPS sequences collected, many had not been identified from the previous HMM iterations (Komatsu *et al.* 2008, Yamada *et al.* 2015, 2016). There were 21 thermophilic class I TPSs, with an extra sequence identified using BLAST, and 8 thermophilic class II diterpene synthases predicted. Only a couple of the thermophilic class II diterpene synthases had neighbouring class I partners with many not containing a class I TPS in the genome. These enzymes might be worth investigating to determine what the resulting product is.

From the predicted thermophilic class I TPSs, six were selected for thermostability and terpene characterisation studies. All six were shown to be mildly thermostable, where only Tchrom and Tbis_3257 were thermostable enough that they might be used in the *Parageobacillus* platform in the future. This study was unable to observe a product from the *T. curvata* TPS, Tcur_3107, through *in vitro* assays where activity was seen in the four other TPSs characterised. While all of the TPSs turned over GPP, they all turned over FPP to a greater extent and were therefore all characterised as sesquiterpene synthases, with the exception of Tbis_3257. JKG1 from *K. papyrolyticum* was shown to be the first example of a hedycaryol synthase from *Chloroflexi* while Tchrom from *T. chromogena* was the first thermostable germacrene D-4-ol synthase to be identified. FJSC11 from *F. thermalis*, although only mildly thermostable, was predicted to be the first (+)-sativene synthase. This terpene has not previously been seen to be produced as a dominant product by a bacterial or plant TPS. Tbis_3257 exhibited activity with both GPP and FPP to form linalool and

farnesene, respectively, but only at a slow turnover rate. However, it was predicted to be a diterpene synthase due to its gene location neighbouring a class II TPS in the *T. bispora* genome. Although it will require further investigation, this TPS could be the first thermostable diterpene synthase produced by a class I/II TPS pair to be characterised. Tbis_3257 was also the first TPS to be characterised with a tryptophan residue replacing the second aspartate residue in the Asp-rich motif. As Tchrom was also thermostable, it should be placed in the *Parageobacillus* terpene system so that a second sesquiterpene can be produced by this system, after τ-muurolol (Styles *et al*., in preparation). The next steps would be to characterise the products of each of these TPSs using NMR and deuterated FPP in combination with GC-MS to confirm their structure and stereochemistry. As this was only a selection of the predicted thermostable class I TPSs, more should be characterised, especially the three predicted TPSs from *A. eurytherma*. As more thermophilic organisms' genomes are sequenced, hopefully more thermostable class I TPSs can be identified.

As very few thermostable class I TPSs have been identified previously and in this work, it would suggest that these enzymes are not very common and extensive work would be required to find more. Regularly run HMMs should be used to increase the chances of finding new sequences for TPSs from thermophiles. One drawback is that this is not a high throughput approach as the products of a TPS cannot be predicted from the sequence or structure of a TPS and therefore characterisation studies are required. Also, there is no dedicated database to thermophiles enabling sequences to be easily identified as being from a thermophile. The number of TPS genes existing in the environment is currently unknown and impossible to estimate as many organisms remain unsequenced or the TPSs are yet to be characterised. However, this work has shown that regular searches for new TPSs is required to sample the ever-increasing number of bacterial genomes being made available. This continued sampling could identify new TPSs and possibly novel terpene structures if continued.

4 Thermostabilisation of a bacterial TPS

4.1 Introduction

Enzymes are biocatalysts that are generally able to recognise and convert substrates to products with high stereo- and regiospecificity. This efficient conversion is attractive for use in industrial processes and is believed to be more sustainable than alternative chemical methods (Singh, Kumar, *et al.* 2017, Chapman *et al.* 2018). Many mesophilic enzymes are used in industry, but some industrial processes benefit from operating under harsh conditions to enable a process to be efficient which prevents mesophilic enzymes from being utilised (Kumar and Satyanarayana 2009, Gurung *et al.* 2013, Jegannathan and Nielsen 2013, Rigoldi *et al.* 2018). This means that stability at extreme temperatures, pHs, salt concentrations or even in organic solvents while maintaining specificity and function is a requirement (Iyer and Ananthanarayan 2008, Stepankova *et al.* 2013). As a result, most common mesophilic proteins cannot be used as they would denature under these conditions and would therefore not function. For high temperature (>50 ^{o}C) related applications, naturally thermostable proteins are utilised from thermophiles, hyperthermophiles and some mesophiles (Iyer and Ananthanarayan 2008, Veith *et al.* 2012, Hussain *et al.* 2013, Rigoldi *et al.* 2018). The advantages of performing a reaction at high temperature include increased substrate solubility, lower risk of contamination when organisms are involved and more importantly, higher rates of reaction (Yu and Huang 2013, Hussein *et al.* 2015, Modarres *et al.* 2016). The most well-known thermostable enzymes used in industry are the DNA polymerases in PCRs from *Thermus aquaticus* and *Pyrococcus* species (Saiki *et al.* 1988, Cline *et al.* 1996). Many of the thermostable enzymes on the market are used to breakdown natural polymers, such as protein, starch and cellulose, to their oligo- and monomeric units (Sarmiento *et al.* 2015, Rigoldi *et al.* 2018). For many processes however, thermostable enzyme variants do not exist or are yet to be discovered. For instance up until 2017, no naturally thermostable TPSs had been characterised (Styles *et al.* 2017). In the absence of natural variants, the thermostability of a mesophilic enzyme that performs the desired reaction can be increased by a number of approaches, including directed evolution, computational methods and prediction software. Several of these methods have been used to increase the thermostability of a TPS.

A successful strategy for increasing the thermostability of enzymes is to alter flexible regions of the protein sequence to form a more rigid structure than the WT at a given temperature (Yu and Huang 2013, Modarres *et al.* 2016). Many computational methods have been developed to identify flexible regions of a protein that can subsequently be rigidified. Some well-known examples are the programs, B-FITTER and FIRST (Floppy inclusion and rigid substructure topography), both of which use the tertiary structure of an enzyme as a starting point (Jacobs *et al.* 2001, Case *et al.* 2005, Reetz and Carballeira 2007, Brooks *et al.* 2009). As with most of the methods used for increasing thermostability, starting with a crystal structure provides more accurate predictions than a homology model or secondary sequence (Yu and Huang 2013). Once flexible regions have been identified in a model or structure, various strategies can be used to increase the thermostability of the protein *in vitro*. Saturation mutagenesis enables residues to be randomly mutated to all other amino acids (Reetz and Carballeira 2006, Gumulya and Reetz 2011). This can be performed on different residues in the flexible region but requires large numbers of mutants to be screened (Reetz and Carballeira 2007). A more informed approach uses multiple sequence alignment of homologous proteins to identify whether there is a consensus in the flexible region (Jochens *et al.* 2010, Anbar *et al.* 2012). If the multiple sequence alignment reveals a different residue or residues in the flexible region to that of the enzyme to be stabilised, these variations can be incorporated into the enzyme (Lehmann *et al.* 2000, Lehmann and Wyss 2001). Alternatively, flexible regions can be rigidified by introducing disulphide bonds, proline residues and salt bridges. While replacing a residue with a proline only requires a single mutation (Takano *et al.* 2009, Tian, Wang, *et al.* 2010), disulphide bonds and salt bridges may require more than one residue to be changed to enable the appropriate bond to form (Badieyan *et al.* 2012, Chen, Yu, *et al.* 2012, Niu *et al.* 2016). In some cases, the packing of the protein has also been shown to be important for stability and mutations to increase the packing of the hydrophobic core of a protein can increase stability (Liu and Kuhlman 2006, Glyakina *et al.* 2007, Chan Joo *et al.* 2011, Mamonova *et al.* 2013). However, changes in amino acids for any of these strategies do not always increase the thermostability of a protein. Our understanding of protein structure has not reached a level that allows prediction of modifications that will definitely increase protein stability especially when a homology model or primary sequence is used. Indeed, introduced modifications regularly have a destabilising effect on the structure or indirectly affect the catalytic rate (Liu and Kuhlman 2006, Johnson *et al.* 2015).

Another method for increasing thermostability is to use prediction software. This relies on either an internal energy function or machine learning to predict thermostabilising mutations (Guerois *et al.* 2002, Capriotti *et al.* 2005, Tian, Wu, *et al.* 2010, Pucci *et al.* 2016). Both normally require a structure as an input, and as mentioned previously a crystal structure, if one is available, is more accurate than using a homology model. In some cases, primary sequences can be used as an input but this is likely to be less reliable than providing a homology model. An example of software that uses an energy function is FoldX which calculates the unfolding free energy of a structure using a variety of parameters. It includes parameter calculations for the van der Waal's and electrostatic interactions between residue side chains and backbone and the entropy score of side chain and backbone conformations (Guerois *et al.* 2002). Some of the parameters are weighted according to a database of mutations from previous studies. Substitutions are simulated at all positions and the unfolding free energy calculated and compared to the WT to find modifications that increase the free energy of unfolding (Guerois *et al.* 2002, Schymkowitz *et al.* 2005). Machine learning programs predict whether a mutation will increase or decrease thermostability without calculating free energy differences between WT and mutant enzymes. These programs use values from a database of how mutations affected the thermodynamics of a protein as well as the experiment conditions. Examples of such databases are Protherm and ProNIT (Abdulla Bava *et al.* 2004, Shaji Kumar *et al.* 2006). I-Mutant and Prethermut are examples of programs that use machine learning to increase thermostability using the Protherm database. The former uses the amino acid sequence while the latter uses the tertiary structure of a protein (Capriotti *et al.* 2005, Tian, Wu, *et al.* 2010). More examples of programs using energy functions or machine learning for improving thermostability are reviewed elsewhere (Modarres *et al.* 2016).

Diaz *et al.* (2011) used the prediction strategy known as SCADS (Statistically, computationally assisted design strategy) to identify residues to increase the thermostability of 5-*epi*-aristolochene synthase (5EAS) from *Nicotiana tobacum* (Starks *et al.* 1997). Using the crystal structure (PDB accession code 5EAT), the SCADS considers the side chain of each residue and scores it on its interaction with neighbouring side chains and the proximal environment around the residue. Overall, structures are scored on their conformational entropy (Calhoun *et al.* 2003). SCADS then determines whether replacing a residue with another amino acid will result in an increased score and therefore a more stable protein. Twelve mutations were predicted and were introduced into 5EAS together. However, it aggregated during expression and had to be refolded prior to characterisation. Using CD, the melt

temperature of the mutant was shown to be 83 °C which was 45 °C above that of the WT. The mutant was also able to catalyse 5-*epi*-aristolochene cyclisation at 65 °C, compared to the WT which was inactivated between 37 °C and 50 °C. Interestingly, even though all of the amino acid changes made were 12 Å away from the active site, the mutant had a >35-fold reduction in turnover compared to the WT and also a reduced product specificity as additional terpene peaks were observed by GC-MS at multiple temperatures. Despite manging to retrieve soluble protein and show a large increase in thermostability of the protein mutant, the distal mutations resulted in the active site cavity being altered. This strategy was clearly successful in identifying thermostabilising mutations but provides a classic example of the inability of a pure computational approach to predict the effects on protein folding. It suggests that changing individual residues should be tested for folding and activity before combinations are attempted.

Directed evolution is a well-known strategy for improving a property of a protein without the need for structural information. It relies on the ability to introduce mutations into a gene at random, with the most common method being to use error-prone PCR (Cobb *et al.* 2013). To enable mutations to be introduced at a defined rate, the PCR protocol is modified by using increased concentrations of magnesium chloride, manganese chloride, uneven nucleotide concentrations or the use of a DNA polymerase lacking the ability to proofread the sequence. Each of these can be used separately or in combination (McCullum *et al.* 2010). Large numbers of mutants are normally screened to identify mutations that enhance a chosen property and, more iterations of error-prone PCR on successful mutations can be used to improve the property further. This method has been utilised to increase the thermostability of many proteins (Kim and Gen Lei 2008, Stephens *et al.* 2009, Mabrouk *et al.* 2013). To maximise the possibility of finding successful mutants from thousands of colonies, a high-throughput screening technique is required. The assays for screening TPSs, including the radioactivity and GC-MS assays, are not high-throughput which makes mutant screens very labour intensive. Lauchli *et al.* (2013) combined the purpald assay of detecting methanol using alcohol oxidase and purpald with terpene cyclisation to create a high-throughput method for mutant screening. Using a vinyl methyl ether substitute for FPP, the cyclisation results in the formation of an unstable hemiacetal which breaks down to release an aldehyde and methanol. The alcohol oxidase converts methanol to formaldehyde which subsequently reacts with purpald to create a purple colour change that can be measured at 550 nm. This enzyme coupled assay was used to screen 2800 mutants of presilphiperforan-8β-ol synthase

(BcBOT2) from *Botrytis cinerea* for thermostabilising mutations. While three successful mutants were obtained from the first iteration, only 19B7 was described, which had both a K85R mutation and a premature stop codon which removed three C-terminal residues. Using the purpald thermostability assay that they developed, 19B7 was shown to retain 50% of its activity after heating at 47 °C over 10 mins, 5 °C above that of the WT. From a second iteration of error-prone PCR on 19B7, 9D6 was identified from a pool of 1800 mutants which additionally contained the H383R mutation. This mutation increased the thermostability further to 54 ^{o}C. These mutations did not affect activity with FPP or the production of presilphiperforan-8β-ol. Whilst, directed evolution has successfully been used to improve enzyme properties, screening thousands of colonies remains labour intensive, even with a high-throughput method. However, unlike the other methods discussed, the screening of the mutants does eliminate any mutations that might reduce folding efficiency or be otherwise destabilising and allows only folded and active mutant proteins to be characterised.

Nine residues in the active site of 5EAS had been identified as relating the active sites of 5EAS from *N. tobacum* and premnaspirodiene synthase from *Hyoscyamus muticus* and that mutation of these residues determined the specificity towards one product or the other(Greenhagen *et al.* 2006, O'Maille *et al.* 2008). Focussing on the same nine residues in the active site of 5EAS, Nartey *et al.* (2017) determined how pairs of mutations affected the unfolding temperature. By mutating two of the residues at a time, 512 mutants were tested for a change in T_m using the Thermofluor assay. The relationship between each individual residue and each pair of residues towards the unfolding temperature was determined using a statistical method called maximum noise entropy. The calculated T_m values show that different pairs of residues could stabilize or destabilize the enzyme and many could not be expressed. The ΔT_m values ranged from -4.1 °C to 11.9 °C compared to the WT. The authors were able to calculate the likelihood of mutations at each position having an effect on 5EAS and whether they were more likely to stabilize or destabilize 5EAS with regards to temperature. This included the synergy between positions even though the positions could be distal in the active site. Despite these mutations forming the active site cavity, mutations at these positions did not significantly affect overall activity, although the product distributions were significantly affected. This work also shows that increasing the T_m does not require an extensive remodelling of a TPS, similar to that performed by Diaz *et al.* (2011), but can be achieved through a small number of mutations.

All of the previous TPS thermostability studies have focussed on the rational engineering and directed evolution of plant TPSs. No studies have been performed with the aim of improving the thermostability of a bacterial TPS. Unlike plant TPSs, there are naturally characterised thermostable τ-muurolol synthases from bacteria (Styles *et al.* 2017). Thus, these can be used as models in an attempt to increase the thermostability of a mesophilic bacterial TPS. To identify more TPSs for use in the *Parageobacillus* terpene platform, this work focussed on finding strategies for increasing the thermostability of bacterial TPSs while maintaining the product specificity and catalytic activity. To try to avoid altering the turnover of the enzyme, mutations were targeted away from the active site cavities and its motifs. Initially, the thermostable τ-muurolol synthases, RoseRS_3509 and Rcas_0662, were compared structurally to the mesostable τ-muurolol synthase, SSCG_03688. As they all produced the same product, these TPSs were likely to have the same catalytic mechanism and similarly structured active sites so that structural differences around the rest of the structure could be compared. Prediction methods were also used to determine thermostabilising mutations from the model of SSCG_03688 and the crystal structure of selinadiene synthase (SdS). These programs were HoTMuSiC, Rosetta and FoldUnfold. The mutants selected by these methods were characterised for differences in terpene product, activity and more importantly thermostability.

4.2 Results and Discussion

4.2.1 Thermostabilisation of SSCG_03688 by sequential removal of the C-terminus

Thermostable and mesostable τ-muurolol synthases from *Roseiflexus* species (RoseRS_3509 and Rcas_0662) and *Streptomyces clavuligerus* (SSCG_03688), respectively, were initially compared by sequence and structural alignments (Figures 4.1 and 4.2). This would allow for direct comparison of the secondary and tertiary structures. As no crystal structures are available for these enzymes, SWISS-MODEL and I-TASSER software were used to build models.

```
Rcas_0662       1   MDQDYRARLVY-PFSGAISPHADIVDQATLAWAAMFGLLTDS-LRHKSRR
RoseRS_3509     1   MDRVDGAQIVY-PFTGAISPYAGDVDQATLIWAEASGLLTDG-LRQKSQR
SSCG_03688      1   MSLNHSDLMFYCPVDDLPHPAASGVNDRTLDWASGQGIPTADRDAGRLRA
                    *.      :.* *. .   * *. *:: ** **   *: * .    : :

Rcas_0662       49  LQYGLLAARAYPRADREMLQIAADWIAWLFFMDDQ-CDET---GIGRDLQ
RoseRS_3509     49  LQYGVLAARAYPQADRDTLQIAADWIAWLFFMDDQ-CDEA---GIGRDPQ
SSCG_03688      51  MAPGLLAARIAPDARGPVLDAFADHHTWLFAFDDEYCDRADGSGITEWAS
                    :  *:****  * *    *:  **  :*** :**: **.:   ** .  .

Rcas_0662       95  RMIALHERFLAILD-GATPEAHDCALTYALADLRRRLALRAPDNWLRRFS
RoseRS_3509     95  RLAALHERFLAVLE-GDPPDSGDWNLTRALADIRRRLAARATDDWLRRFG
SSCG_03688      101 FLARLH----RVVETGESALLPGNPYGLALRDIACRLSTYTTPAQLAEWL
                     :  **     ::: * ..   .     ** *:  **:  :.   * .:

Rcas_0662       144 EHVRLYFTANRWETVNRQ-RGATPNVATYCAARLFSGAVYACFDLIELAE
RoseRS_3509     144 EHVRLYFTANRWEAANRC-QSIVPNVATYCAARLFSGAVYACFDLIELAA
SSCG_03688      147 EALRSYFAALVWERSRRRDDDRLQSLDDYLLLRLRNGAMHTSITLLDTVN
                    * :* **:*  **  .*   .   .:  *   ** .**:::.: *:: .

Rcas_0662       193 QIELPFYARHHSIVQQLEQAANNIICWCNDVLSYPKEMQHGDRH-NLVLV
RoseRS_3509     193 GIDLPFYARYHAAVQQLERTANNIICWCNDMLSYPKEMQHGDVH-NLVLA
SSCG_03688      197 GYVLPRELRETPGVRALVEMTALLVSVDNDILSHHKESTSGTREANLLDV
                       **   *  . *: * . :  ::.  **:**: **   *  . **: .

Rcas_0662       242 IQGEHQCSLPEAIDRALDLHAREVATFVRKRTCVPYFDAAVNTALEKYVT
RoseRS_3509     242 IRQEYQCSLSEALQQALLLHDRETDTFMKTQKQLPRFNPAVDMALERYID
SSCG_03688      247 LGRTGHTTPGEAVAQAVALRNEIMRQFVRVAERVR--TPAAVPELYRFTT
                    :    : :  **: :*: *: .    *::    :    .*.   * ::

Rcas_0662       292 GLQFWICANRDWSLTATRYAPT----------------------------
RoseRS_3509     292 GLQYWICANRDWSLTAMRYALS----------------------------
SSCG_03688      295 GLARWIRANLDFSLTTTRYTGPVTERAALSPHEVPPLSGQGPAPARSDVI
                    **  ** ** *:***: **: .

Rcas_0662       314 ----------------HK-------------------------SQEMV--
RoseRS_3509     314 ----------------GT-------------------------EPVLR--
SSCG_03688      345 GWWWRIPEPLPEPGSDGADTPVRKRRAGDRPPTAGRGGAPHHQRTGPPPP


Rcas_0662       321 ---MAVA---------------QQ
RoseRS_3509     321 ---TRFA---------------HS
SSCG_03688      395 VLPGGITASRSSGLQQSTWRREHR
                        .:              :
```

Figure 4.1: Sequence alignment of the τ-muurolol synthase amino acid sequences of RoseRS_3509, Rcas_0662 and SSCG_03688. The sequences labelled in red are the Asp-rich, pyrophosphate sensor, NSE triads and RY motifs in that order. This alignment was made using the multiple sequence alignment tool, T-coffee.

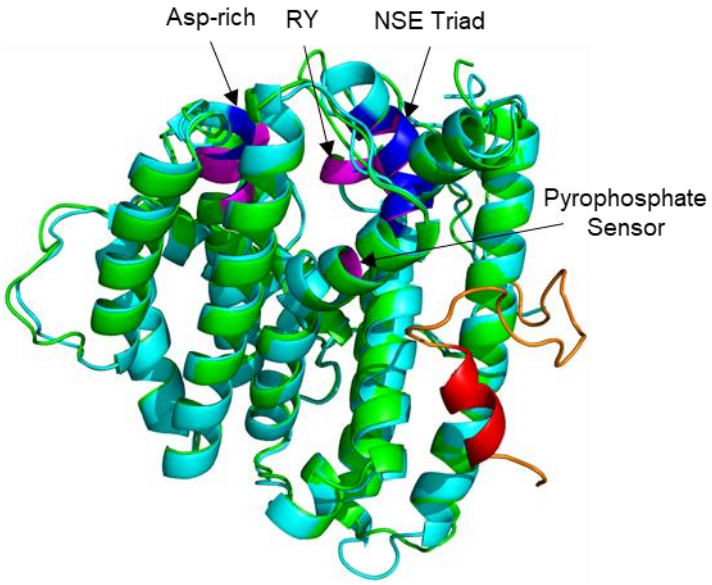


Figure 4.2: Aligned I-TASSER models of RoseRS_3509 (Cyan) and SSCG_03688 (Green). The extended C-terminus of SSCG_03688 is labelled in orange and the modelled C-terminus α-helix of SSCG_03688 labelled in red. The C-terminal was not modelled from P353 onwards as no consistent position was predicted. The Asp-rich, NSE triad, pyrophosphate sensor and RY motifs of RoseRS_3509 and SSCG_03688 are shown in blue and magenta, respectively.

Although the sequence identity between RoseRS_3509/Rcas_0662 and SSCG_03688 was < 24%, the sequences and structures show good alignments including the conserved motifs; Asp-rich, pyrophosphate sensor, NSE triad and RY motifs. The only clear structural difference was the extended C-terminal of SSCG_03688 from P330 onwards. This tail extended 92 residues further than the thermostable variants, with the only secondary structure that was consistently predicted by both SWISS-MODEL and I-TASSER being a small α-helix of different lengths with the core residues W346, W347, W348 and R349 (Figure 4.2). The function of this extension is unknown as no proteins containing a similar extension were identified when the residues from P330 onwards were run through BLAST. After this helix, the various models of SSCG_03688 could not consistently predict the position of the rest of the C-terminal leaving no indication as to whether it is highly mobile or forms an integral part of the core structure. Previously, the removal of three residues from the C-terminal of BcBOT2 combined with the K85R mutation increased the T_m by 5 °C (Lauchli *et al.* 2013) but it was not determined whether one or both of the mutations was responsible for increasing the thermostability of BcBOT2. To determine whether the removal of the extra residues of SSCG_03688 would increase its thermostability or affect function, blocks of 8-14 residues were sequentially cut

from the protein sequence (Figure 4.3). The most truncated mutant was a similar length to RoseRS_3509 and Rcas_0662 with a total of 90 residues removed.

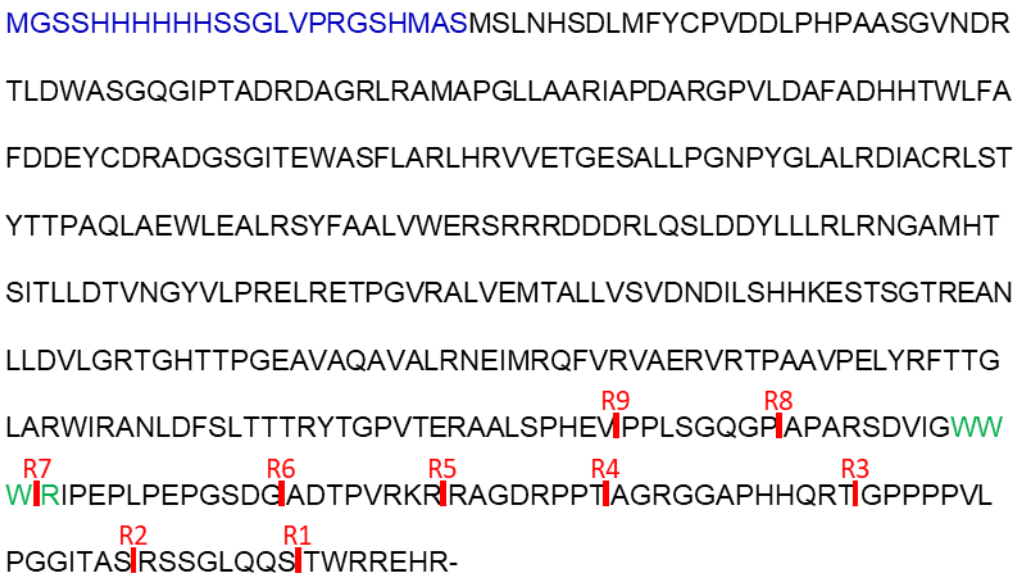


Figure 4.3: C-terminal truncations of SSCG_03688. The N-terminal His-tag (Blue), the C-terminal α-helix (Green) and the positions where the TAA stop codons were introduced to the C-terminal of SSCG_03688 were also labelled. Truncations were named SSCG_R1 to R9.

Each truncation was made by PCR amplification of the gene with the reverse primer introducing a stop codon to specifically amplify a shortened gene. The least truncated mutant with seven residues removed was SSCG_R1 and the most truncated, SSCG_R9, had 89 residues removed (Figure 4.3). The primers added NheI and EcoRI restriction sites for restriction and ligation into pET28a. All truncated genes were inserted successfully into pET28a and transformed into *E. coli* Bio*Blue* and *E. coli* BL21(DE3) (Figure 4.4A). The same expression conditions used for *sscg_03688* by Hu *et al.* (2011) were used here for *sscg_03688* and its truncation mutants. Soluble protein was obtained for SSCG_R1-6 (Figure 4.4B), however, under the same conditions, SSCG_R7-9 formed insoluble aggregates (Figure 4.4C).

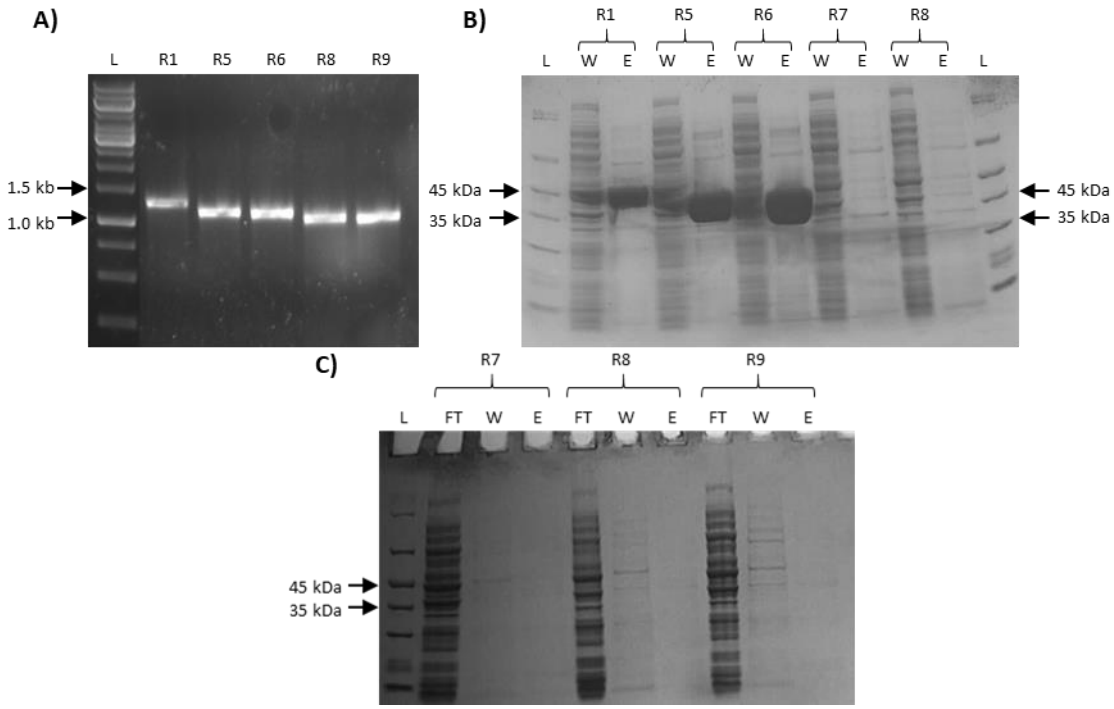


Figure 4.4: Agarose gel showing the PCR amplification of *sscg_03688 truncation* mutant genes and SDS-PAGE gels showing the affinity purification of the SSCG_03688 truncation mutants. A) The PCR amplified *sscg_R1* (~1.3 kb), *sscg_R5* (~1.2 kb), *sscg_R6* (~1.2 kb), *sscg_R8* (~1.1 kb) and *sscg_R9* (~1.1 kb) truncated genes restriction digested for ligation into pET28a, analysed on a 1% agarose gel. B) An example SDS-PAGE gel of the soluble fractions from the affinity purification of SSCG_R1 (47.4 kDa), R5 (43.3 kDa), R6 (42.3 kDa), R7 (41.0 kDa) and R8 (39.6 kDa) through the purification process. C) Soluble fraction of the repeat purifications of SSCG_R7, R8 and R9 (38.8 kDa) did not produce any soluble protein. All of the SSCG truncation mutants were purified in 1L of *E. coli* BL21(DE3) culture. L – the ladder used for the DNA gel was GeneRuler 1kb DNA Ladder and for the SDS-PAGE gels was Unstained Protein Molecular Weight Marker, FT – Flow through, W – Buffer Wash, E – Elution. FT, W and E were the fractions from the affinity chromatography purification.

Prior to induction of 500 ml *sscg_R7-R9* expression cultures with IPTG, a 1 ml sample was taken and the genes were PCR amplified and sequenced. It was identified that the *E. coli* BL21(DE3) strains used contained the pET28a:*sscg_R7-R9* constructs with 100 % identity to the reference sequence. This indicated that the presence of the *sscg_R7-R9* genes were not toxic to growth as the cultures grew well and therefore the complication lay post-induction. This meant that only 57 residues could be removed successfully before protein folding/stability was compromised. Compared to SSCG_R6, the further truncation mutants, SSCG_R7, R8 and R9 were meant to remove a further 13, 25 and 33 residues, respectively. However, as the 13 residues removed in SSCG_R7 are common to all three truncations, it is likely that the loss of these is the cause of the structural instability and are required for soluble

SSCG_03688 to be produced. Although the general position of most of the residues cleaved by SSCG_R7 could not be accurately predicted by modelling, the final residues, W350, W351 and W352, were modelled to form part of the small C-terminal α-helix with R349 as previously mentioned. This helix is observed in the crystal structure of selina-3,7(11)-diene synthase (PDB accession code 4OKM) from *Streptomyces pristinaespiralis* where it is made up of the three tryptophan residues with an aspartate replacing the arginine present in SSCG_03688 (Baer, Rabe, Fischer, *et al.* 2014). Also, the Cembrene C synthase sequence from *Streptomyces* sp. SANK 60404 and ROS37996 from *A. eurytherma*, identified in chapter 3, contained this triple tryptophan followed by a glutamate or aspartate, respectively (Meguro *et al.* 2013). This suggests that amongst these enzymes, this sequence and the α-helix is conserved. Between this helix and the C-terminus, SSCG_03688 extends for another 69 residues where the other sequences have between 2 to 19 residues. Most other TPSs have shorter C-terminal sequences and are therefore missing the α-helix. The variability of sequence length suggests that the removal of any residues involved in the α-helix, including R349, was likely to be the cause of the instability rather than the other residues that could not be modelled. As this helix is translated later than the rest of the structure, its absence might expose a hydrophobic region or prevent key salt bridges forming resulting in protein aggregation. For further C-terminal truncation to be continued, stabilising mutations would be required to compensate for the loss of this helix. Sequential truncation of individual residues between SSCG_R6 and R7 would be necessary to determine the maximum number of residues that can be removed before stability is affected and therefore which residues are vital for folding. This could be compared to sequential truncation of SdS, Cembrene C and ROS37996 to show whether this helix is vital for protein stability in these enzymes as well. Without a crystal structure and further study of SSCG_03688 and SSCG_R6, this remains a conjecture. However, as no soluble protein was retrieved for SSCG_R7-9, these mutants were not investigated further.

Initially, to determine whether the C-terminal truncation affects thermostability, wavelength scans and thermal melts using CD were performed to compare the secondary structure and melting temperatures of SSCG_03688 and a selection of the truncated mutants (Figures 4.5 and 4.6). The wavelength scans were performed between 200-300 nm. The scans showed that the WT and the soluble truncated mutants retained the α-helical structure of class I TPSs due to the presence of the double minima around 208 nm and 222 nm (Diaz *et al.* 2011, Lauchli *et al.* 2013). The T_m values calculated from the thermal melt curves showed that there was an overall

increase of ~1.2 °C between SSCG_03688 and SSCG_R6, which was the most thermostable mutant. Removal of the first seven residues to form SSCG_R1, did not increase the T_m according to the melt curves, however, removal of 41 residues or more resulted in an increase of ~0.85 ^oC. This suggested that the more the C-terminal of SSCG_03688 was truncated, the higher the T_m (Table 4.1). This was also shown by the small increase in the van't hoff enthalpy (ΔH) of SSCG R6 compared to SSCG_03688. To identify whether this is significant change, replicates would be required. As SSCG_R6 was the most thermostable mutant, it was the only mutant considered for further experiments.

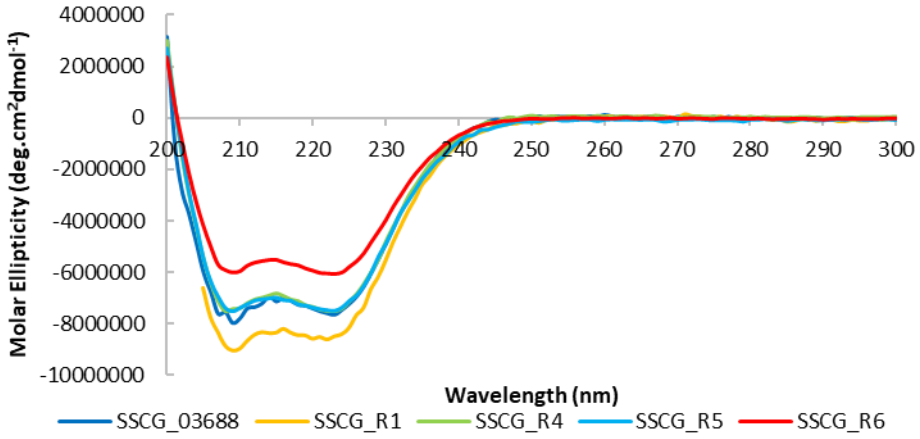


Figure 4.5: Wavelength scans of SSCG_03688, SSCG_R1, R4, R5 and R6 at 20 °C to determine each proteins secondary structure using CD. Wavelength scans were performed between 200-300 nm except for SSCG_R1. Due to the lower dilution in CD buffer, SSCG R1 could only be run to 205 nm to avoid the noise produced at lower wavelengths that would lower the reliability of the data.

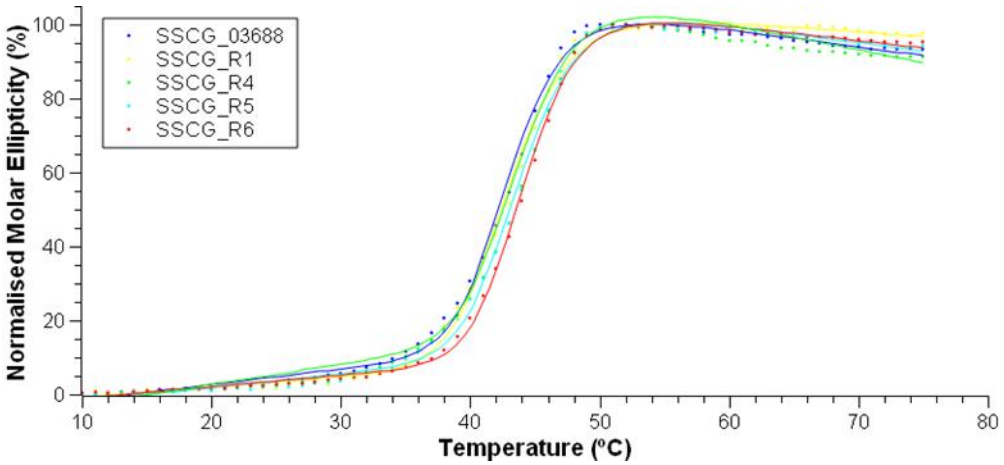


Figure 4.6: Normalised molar ellipticity of the thermal melts of SSCG_03688, SCCG_R1, SSCG_R4, SSCG_R5 and SSCG_R6 using CD at 222 nm between 10-75 °C to determine the melt temperatures. The R^2 values were calculated as 0.999.

Table 4.1: Melting temperatures of SSCG_03688 and its truncated mutants calculated from the curves in Figure 4.6 from a single replicate. The error is the deviation of the data from the curve created by Equation 4 (Chapter 2.6.5).

Terpene Synthase	T_m (°C)	ΔH (kJ/mol)
SSCG_03688	43.06 ± 0.07	45.0 ± 1.4
SSCG_R1	43.06 ± 0.06	41.4 ± 1.0
SSCG_R4	43.90 ± 0.07	45.0 ± 1.2
SSCG_R5	43.95 ± 0.08	44.0 ± 0.9
SSCG_R6	44.25 ± 0.05	46.9 ± 0.8

As a second method to compare the thermostability of SSCG_03688 and SSCG_R6, irreversible denaturation assays were carried out using GC-FID to measure residual activity for product quantification as a measure of thermostability encompassing both structure and activity. This method enables the measurement of the rate at which an enzyme moves from either active or reversibly inactive states to an irreversible inactive state. The rate of inactivation was calculated using equations 2 and 3 in chapter 2.6.4.4. As the concentration of active and reversibly inactive protein is proportional to the enzyme rate (under substrate saturation), the catalytic rates of each sample can be compared to the zero time-point to determine the rate of inactivation. Initial inactivation experiments showed that SSCG_03688 steadily lost activity over the course of an hour at 35 °C, where inactivation assays at 40 °C and above resulted in complete inactivation after ~30 mins (Data not shown). For all

further inactivation studies for SSCG_03688 and SSCG_R6, 35 °C was therefore used. The decline of activity after incubation at 35 °C of SSCG_03688 and SSCG_R6 over the course of an hour were compared (Figure 4.7 and Table 4.2).

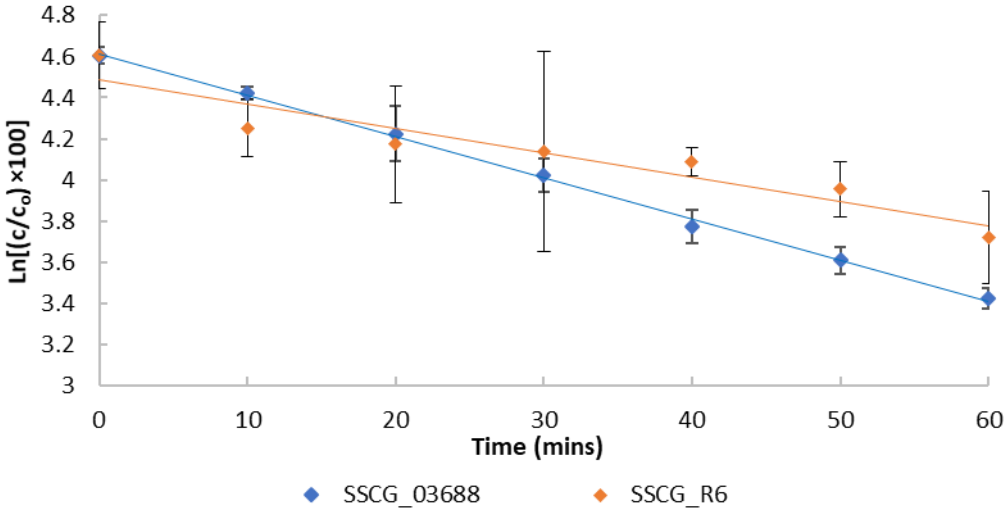


Figure 4.7: Inactivation of SSCG_03688 and SSCG_R6 at 35 °C using the irreversible thermal denaturation assay. The inactivation rates were calculated as the gradients of the linear fits. R^2 values for the linear fits were 0.998 and 0.897 for SSCG_03688 and SSCG_R6, respectively. Assays were performed in triplicate.

Table 4.2: Inactivation rates calculated for SSCG_03688 and SSCG_R6 from the negative gradient of the plots in Figure 4.7.

Terpene Synthase	Inactivation Rate (min^{-1})
SSCG_03688	0.0201
SSCG_R6	0.0119

Both enzymes showed a steady decrease in activity when heated at 35 °C (Figure 4.7). However, when the rate of irreversible inactivation was determined, it showed that SSCG_R6 was inactivated more slowly than SSCG_03688 at this temperature indicating that it is more stable than SSCG_03688 at 35 °C (Table 4.2). In conjunction with the data from the CD thermal melts, this suggests that the C-terminal truncation had a stabilising effect. It also might suggest that the significant increase in thermostability of BcBOT2 during previous work was at least partly due to the K85R mutation as well as the three-residue truncation at the C-terminal (Lauchli *et al.* 2013).

The C-termini of most bacterial TPSs do not extend nearly far enough to associate with the active site, however the C-terminal extension of SSCG_03688 could potentially be an exception. The C-terminal extension of the most similar length comes from SdS, which, from the crystal structures, clearly does not interact directly with the active site (Baer, Rabe, Fischer, *et al.* 2014). On the other hand, the C-terminal extension of SSCG_03688 could extend round to the active site. A possible function of this extension could be to cover the active site for closure and solvent exclusion or interact with the active site and aid Mg^{2+} ion or substrate binding. A crystal structure of SSCG_03688 in a closed conformation might resolve this problem by showing the C-terminal in close proximity to the active site. However, in the absence of a crystal structure, GC-MS assays were used to determine if the removal of the C-terminus might affect the product distribution profile. GC-MS assays were performed in triplicate with SSCG_03688 and SSCG_R6 at 30 °C and 40 °C to directly compare the product distributions at various temperatures (Figures 4.8).

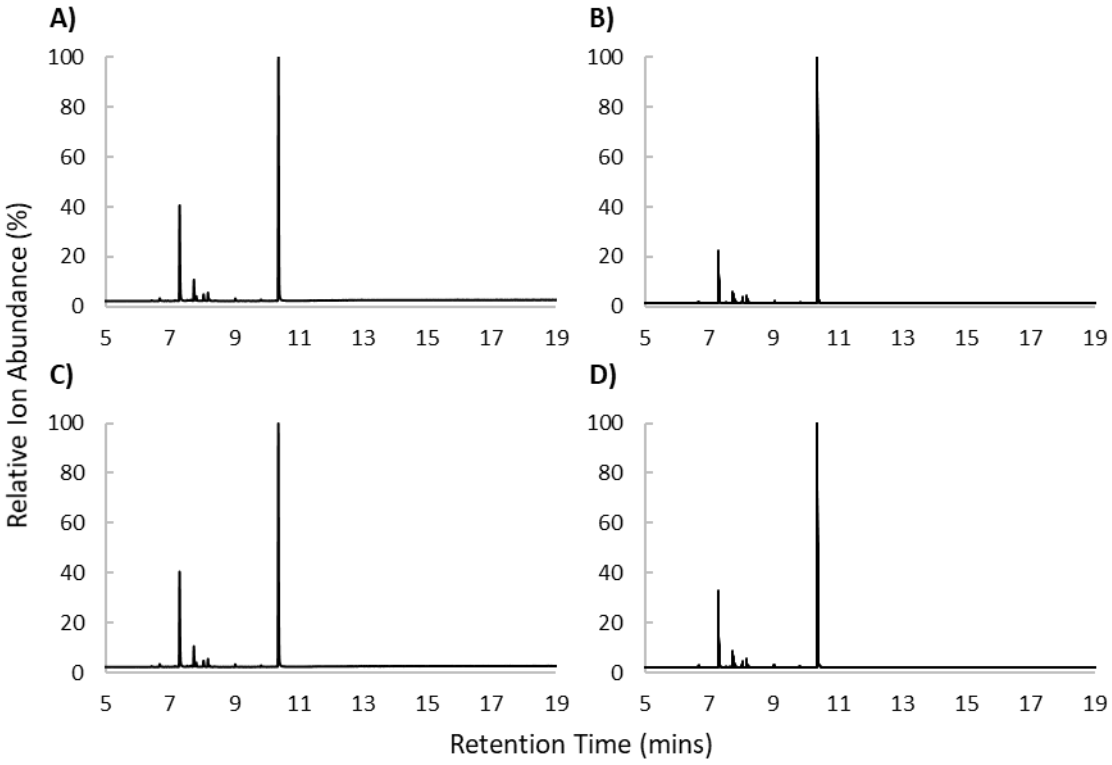


Figure 4.8: Normalised GC-MS spectra of SSCG_03688 at A) 30 °C and B) 40 °C and SSCG_R6 at C) 30 °C and D) 40 ºC. The dominant peak at 10.367 mins was τ-muurolol with another minor product at 7.212 mins.

All four spectra show τ-muurolol as the dominant product with several minor products with shorter retention times also evident. This shows that the removal of the C-terminal does not seem to alter the product distribution significantly even at varying temperatures suggesting that it does not associate with the active site and partial removal leaves the cyclisation mechanism towards τ-muurolol formation intact. To determine whether the truncation indirectly affected the steady-state kinetics of the enzyme, GC-FID assays at varying FPP concentrations with 1 µM SSCG_03688 and SSCG_R6 were performed at 30 °C (Table 4.3).

Table 4.3: Steady state kinetic parameters of SSCG_03688 and SSCG_R6 calculated using the GC-FID assay. The concentrations of FPP were between 0-92.4 µM. The 46.2 µM data point was removed for SSCG_R6 as it was an erroneous result.

Terpene Synthase	**K_m (µM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_m ($\times 10^3$ $s^{-1}M^{-1}$)**
SSCG_03688	4.5 ± 1.0	0.018 ± 0.001	4.0 ± 0.9
SSCG_R6	16.3 ± 4.0	0.083 ± 0.006	5.1 ± 1.3

The kinetics of SSCG_03688 were originally characterised by Hu *et al.* (2011) using the radioactive assay. Hu *et al.* (2011) estimated the K_m, k_{cat} and k_{cat}/K_m values to be 2.7 µM, 0.0016 s^{-1} and 6.04 × 10^2 $s^{-1}M^{-1}$, respectively. These values are lower than those calculated using the GC-FID assay in this study. Vardakou *et al.* (2014) showed that although the values for the radioactive assay are expected to be lower for K_m and k_{cat} compared to GC-MS, the catalytic efficiency was expected to be relatively similar between the two methods. Surprisingly, the catalytic efficiency was over 6-fold lower based on the radioactive assay compared to the GC-MS assay. When SSCG_R6 was compared to SSCG_03688, both the K_m and the turnover number increased by a similar proportion. This resulted in similar catalytic efficiencies of SSCG_03688 and SSCG_R6. This suggests that the C-terminal truncation does not have a significant effect on the catalytic efficiency or τ-muurolol cyclisation mechanism. Overall, the C-terminal truncation strategy may be useful to gain small increases in thermostability of proteins with an extended C-terminal, including TPSs. The C-terminal is likely to be a mobile component with many conformations and therefore its removal prevents unfolding by its increased motion under higher temperatures. As the presence of the helix seems key to stability probably through interaction with neighbouring helices, the B-factor of the rest of the C-terminus will be lower than the section that was removed. As the complex interactions that enable Rcas_0662 and RoseRS_3509 to

be thermostable cannot be truly determined without more structural information, it would be difficult to predict single or multiple mutations to increase the thermostability of SSCG_R6 through comparison of homology models. To increase the thermostability of SSCG_R6 further, mutations are probably required in its core structure, which may be better predicted by using various computational software.

4.2.2 Computational predictions for thermostabilisation of terpene synthases

To try to increase the thermostability of SSCG_R6 further, the prediction program, HoTMuSiC was used (Pucci *et al.* 2016). This program aims to increase the T_m of an enzyme by performing *in silico* residue substitutions that are scanned for increases in the unfolding free energy. The I-TASSER models of SSCG_03688 and SSCG_R6 were input into the server with their T_m values as calculated using CD. The program then systematically ran through all of the residue positions and changed each residue to the other 19 aa in turn to determine which mutations might increase the T_m. Of the total 6853 mutations tested for SSCG_R6, 913 were believed to increase the T_m. The full list of mutations predicted for SSCG_03688 and SSCG_R6 by HoTMuSiC is attached supplementary media under the filename “SSCG_R6 HoTMuSiC Mutation Prediction”. Any mutations within the active site cavity or the conserved motifs were not considered for modification to prevent catalysis or the mechanism from being affected. To narrow the number of mutations down further, only mutations that were predicted to increase the T_m in both SSCG_03688 and SSCG_R6 were considered. The positions of the predicted mutations were also compared to the same positions in RoseRS_3509 and Rcas_0662. Of the remaining mutations, seven were selected for further investigation. These were G24I, P39I, S100Q, A115P, G209A, D224C and T293L (Figure 4.9). Of these residues, most were positioned on α-helices with two mutations, P39I and A115P being on loops. The majority of the residues were predicted by the model to be solvent exposed except for T293 which was predicted to be buried at the base of the protein away from the active site. Also, D224 is adjacent to the start of the NSE triad sequence, although it does not line the active site cavity. It was uncertain whether mutating this residue would result in catalysis being affected.

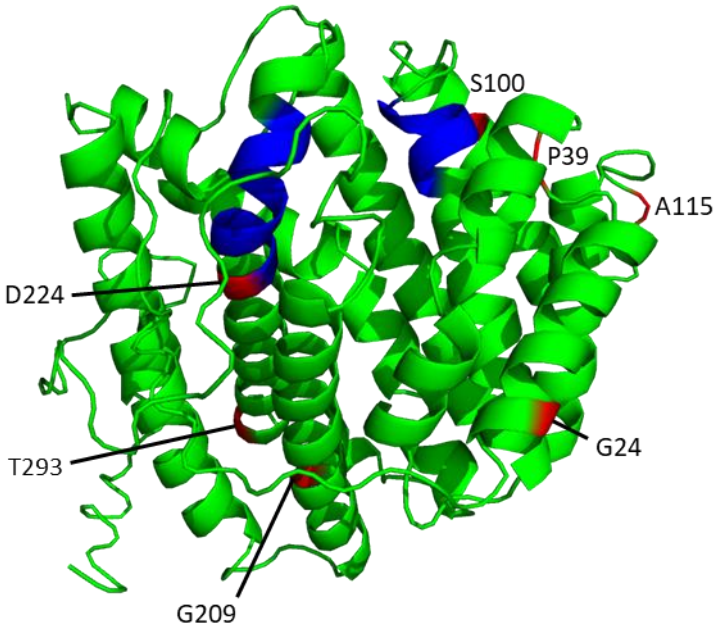


Figure 4.9: Positions of the mutations on the I-TASSER model of SSCG_R6 predicted by HotMuSiC. Each mutation is show in red and the motifs are shown in blue with the NSE triad on the left and Asp-rich on the right of the model.

Each mutation was introduced into *sscg_R6* by overlapping PCR and the mutant gene inserted into the pET28a plasmid through restriction digestion and ligation between the NheI and EcoRI restriction sites. Each construct was transformed into *E. coli* Bio*Blue* and *E. coli* BL21(DE3) strains and their sequences confirmed prior to expression studies. The same conditions that were used for *sscg_03688* and *sscg_R6* expression were used for the mutants (Hu *et al.* 2011). SDS-PAGE gels of the affinity purifications of each mutant are shown in Appendix Figure 12.

Large quantities of soluble G24I, S100Q and G209A were obtained from affinity purification according to the SDS-PAGE gels with some protein in the insoluble affinity purification fractions. SDS-PAGE gels of soluble and insoluble fractions from expression of the A115P and D224C mutants suggested that these proteins aggregated due to mis-folding during expression. Most of the T293L mutant protein also appeared in the insoluble cell debris fraction after expression, although some soluble protein was recovered from affinity purification. However, while a good quantity of soluble P39I was eluted from the affinity purification step, the SDS-PAGE gel suggested that this protein was unstable as intense bands were observed in the insoluble fractions of the purification. This was confirmed when P39I precipitated out in the eluate as well as when it was buffer exchanged and concentrated. Similarly,

the T293L mutant protein also precipitated when concentrated. This meant that little to no P39I and T293L could be isolated and indicated that these mutations, including A115P and D224C, had a destabilising effect on the protein. This might be as a result of HoTMuSiC only being supplied with a model rather than a crystal structure, although mutations predicted using similar programs do not always increase thermostability. The model is only a prediction of the position of each residue and their interactions towards stability of the protein. This would allow mutations to be predicted that destabilise an area of the protein rather than stabilise. Without a crystal structure, it would be difficult to determine what interactions are possibly disrupted to result in these mutations causing instability. Unlike in Diaz *et al.* (2011), no refolding experiments were performed as the mutants which aggregated would not be useful for any downstream experiments without the introduction of stabilising mutations. Therefore, as G24I, S100Q and G209A were the only mutants that naturally produced soluble proteins, these were taken forward for further study.

Initially, the secondary structures and melting temperatures were measured using CD (Figures 4.10 and 4.11). The mutations did not perturb the α-helical structure, with all mutants showing minima at 208 nm and 222 nm for the wavelength scans similarly to SSCG_R6.

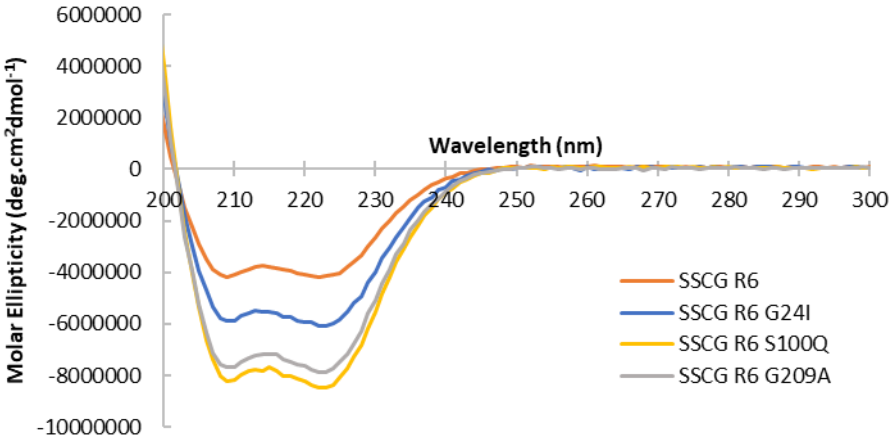


Figure 4.10: CD wavelength scans to determine the secondary structures of SSCG_R6, G24I, S100Q and G209A between 200-300 nm performed at 20 ºC.

As the secondary structures were as predicted, SSCG_R6 and the mutants were melted between 20-70 °C to calculate the T_m and van't Hoff enthalpy (ΔH) using fresh CD buffer (Table 4.4). While the T_m for SSCG_R6 was calculated as 45.5 ºC, G209A had a similar T_m as 45.4 °C, and G24I showed an increase in the T_m to 46.6 °C. While

the replacement of G24I seemed to have a stabilising effect, the S100Q mutant, however showed a decrease in the T_m to 43.2 °C suggesting that it is destabilizing the structure. The ΔH values suggest that the unfolding of S100Q and G209A is much more thermodynamically favourable than SSCG_03688 while unfolding of G24I was less favourable. As fresh CD buffer was made for this experiment, only the T_m values for SSCG_R6 and its mutants described here can be compared to one another. The mutation of glycine at position 24 may increase thermostability by reducing the number of conformations that can be accessed and therefore lower the flexibility of the protein backbone. This would result in a decrease in entropic contribution to the unfolding free energy (Tian, Wang, *et al.* 2010, Anbar *et al.* 2012, Torpenholt *et al.* 2015, Veno *et al.* 2017). It should be noted that in many cases, the introduction of glycine can increase the flexibility by removing constraints on the backbone resulting in an increase of thermostability (Anbar *et al.* 2010, Yi *et al.* 2010). The mutation of S100 to glutamine may disrupt key interactions on the surface of the α-helix, therefore lowering its stability. Crystal structures of SSCG_03688 and SSCG_R6 would aid in determining how these mutations affect interactions around these residues resulting in an increase or decrease in thermostability.

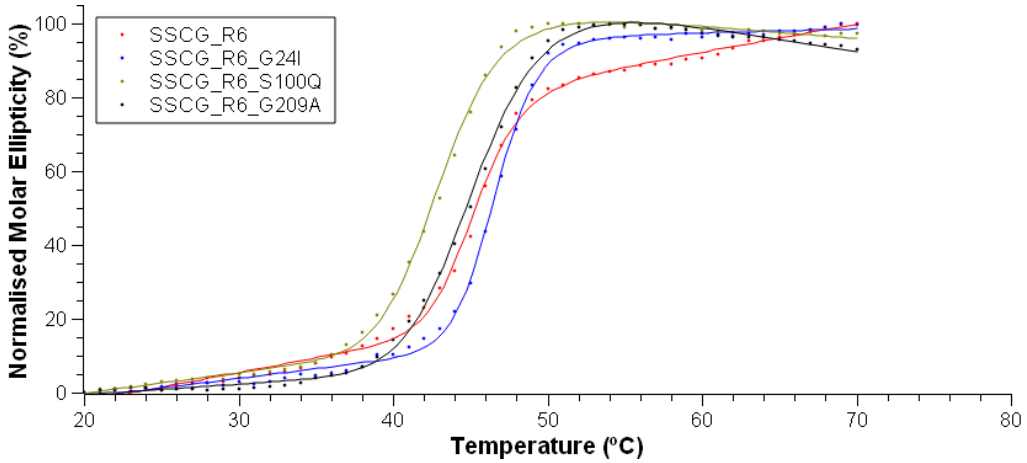


Figure 4.11: Thermal melt curves of SSCG_R6, G24I, S100Q and G209A. The change in molar ellipticity was followed at 222 nm as the temperature was increased from 20-70 °C using CD. The R^2 values were 0.999 for all curves.

Table 4.4: T_m and van't Hoff enathalpy (ΔH) values for SSCG_R6 and the single point mutants, G24I, S100Q and G209A.

Terpene Synthase	T_m (°C)	ΔH (kJ/mol)
SSCG_R6	45.26 ± 0.09	62.7 ± 2.7
SSCG_R6_G24I	46.65 ± 0.04	73.0 ± 2.2
SSCG_R6_S100Q	43.20 ± 0.05	48.4 ± 1.0
SSCG_R6_G209A	45.46 ± 0.05	48.1 ± 1.1

To determine the effect of these changes on irreversible thermal denaturation assays were performed on these mutants as in chapter 4.2.1. The mutants were heated at 35 °C, with samples being removed every ten minutes to assay for activity. The terpene products were extracted using a hexane overlayer and analysed by GC-FID for product quantification to enable the rates of inactivation for each mutant to be calculated.

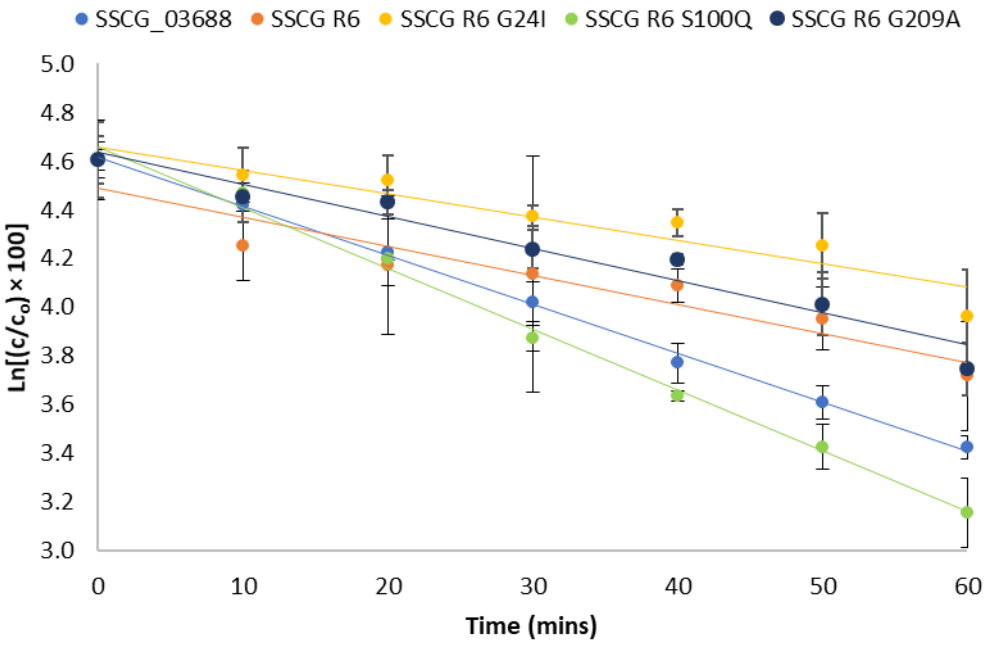


Figure 4.12: Irreversible thermal denaturation assays of SSCG_R6 and its mutants when heated at 35 °C. Linear regression was performed and the R^2 values were calculated as 0.891, 0.994 and 0.950 for G24I, S100Q and G209A, respectively. Assays were performed in triplicate. The SSCG_03688 and SSCG_R6 data are the same as in Figure 4.7.

Table 4.5: Inactivation rates of G24I, S100Q and G209A at 35 °C. The values for SSCG_03688 and SSCG_R6 are from Table 4.2.

Terpene Synthase	Inactivation Rate (min^{-1})
SSCG_03688	0.0201
SSCG_R6	0.0119
SSCG_R6_G24I	0.0096
SSCG_R6_S100Q	0.0250
SSCG_R6_G209A	0.0132

SSCG_R6_G24I had the slowest rate of inactivation at 35 °C while S100Q had the highest. This suggests that S100Q is catalytically inactivated faster than SSCG_03688 but also a less stable secondary structure according to the thermal melt data. SSCG_R6_G209A had a marginally higher inactivation rate than SSCG_R6 suggesting that although the rate of secondary structure unfolding is relatively similar between SSCG_R6 and SSCG_R6_G209A, as shown by the T_m values, the active site is being compromised faster as shown by the higher inactivation rate. The inactivation rate, does however, remain below that of SSCG_03688. The only mutant to have a more stable secondary structure and a slower rate of irreversible denaturation was G24I.

Even though these mutations were not near to the active site, the kinetics of each enzyme were determined using the malachite green assay at 30 °C, to identify any indirect effects.

Table 4.6: Kinetic constants for G24I, S100Q and G209A calculated using the malachite green assay at 30 °C. The error is shown as standard deviation. Kinetic assays were performed with FPP concentrations between 0-115.4 µM with outliers removed. The values removed were 23.1 µM for G24I and 11.6 µM and 115.4 µM for S100Q and G209A. Michaelis-Menten kinetic curves are shown in Appendix Figure 17. Errors shown are standard errors. *The kinetic parameters for SSCG_R6 were calculated using the GC-FID assay and are also shown in Table 4.3.

Terpene Synthase	K_M (µM)	k_{cat} (s^{-1})	k_{cat}/K_M ($x10^{-3}$ $s^{-1}mM^{-1}$)
SSCG_R6*	16.3 ± 4.0	0.083 ± 0.006	5.1 ± 1.3
SSCG_R6_G24I	3.8 ± 0.8	0.039 ± 0.002	10.3 ± 2.2
SSCG_R6_S100Q	7.9 ± 0.9	0.054 ± 0.002	6.8 ± 0.8
SSCG_R6_G209A	3.0 ± 0.8	0.026 ± 0.001	8.5 ± 2.2

Both of the more structurally stable mutants had lower K_M values than SSCG_R6_S100Q, but also had lower k_{cat} values. Overall, the catalytic efficiencies suggested that SSCG_R6_G24I was the most efficient of the mutants and SSCG_R6_S100Q was the least efficient. This suggests that as well as altering the stability of the protein, these residues indirectly interact with the active site and therefore the catalytic mechanism. Although the kinetic parameters for SSCG_R6 were determined using the GC-FID assay, the catalytic efficiencies should still be comparable to the malachite green assay (Vardakou *et al.* 2014). The catalytic efficiencies suggest that these mutations do not disrupt catalysis or the product distribution of these enzymes. Indeed, the kinetic parameters suggest that these mutations may have a positive effect on catalytic efficiency.

As the first attempt to use a program to predict substitutions which increased thermostability only identified a single stabilising mutation when applied to a model structure, the next attempt included a TPS with a crystal structure. Alongside the SSCG_R6 I-TASSER model, the tetrameric crystal structure of SdS (PDB accession code 4OKM) (Baer, Rabe, Fischer, *et al.* 2014), was used to try and identify thermostabilising mutations. For this, a combination of programs including the point mutant (pMut) scan application from Rosetta and the online program, FoldUnfold, were used (Galzitskaya *et al.* 2006, Kellogg *et al.* 2010). The pMut script looks for mutations that increase the stability through the internal scoring system used by the Rosetta database. To calculate which residue substitutions to make, Rosetta scripts use their own scoring and energy function taking into account hydrogen bonds, backbone and side chain conformations, solvation and more. The energy function is described in detail in Alford *et al.* (2017). The FoldUnfold server allows any protein sequence to be input and it will calculate the change in packing density across the sequence using a trained dataset. This enables disordered and folded regions of a protein to be predicted even if the precise structure of a protein is not known. It also provides a mean score to the overall structure in terms of packing density. The higher the value, the more folded the protein is or the smaller the proportion of unfolded structure (Galzitskaya *et al.* 2006). This score would allow for the packing densities of potential mutants to be compared to the WT. As SSCG_R6, did not have a crystal structure, it was hoped that using FoldUnfold would aid decisions for subsequent site-directed mutagenesis by enabling flexible regions to be targeted. Before running these sequences through the pMut script and FoldUnfold, the ligands and water molecules from the structures were cleaned/removed using the clean PDB python script from Rosetta so that the structures could be used in the pMut Rosetta script

that does not tolerate extra moieties. The structures were then relaxed using the relax script from Rosetta. This enables the optimal positions of side chains and the protein backbone to be estimated and the local energy minima for the structures to be identified. This helps prepare the structure for the pMut script. As introducing multiple mutations at once tends to increase the chances of instability or loss of function rather than stability of a protein, single point mutations were targeted. To do this, the Rosetta pMut script was used on the relaxed structures of SSCG_R6 and SdS to predict stabilising mutations. Of the possible 6859 mutations tested for SSCG_R6, only 108 were predicted by the pMut script to be stabilising. For the SdS tetramer this was 131 of 25707 potential mutations. The full list of mutations predicted by pMut for SSCG_R6 and SdS are on the attached supplementary media under the filename "Rosetta pMut Mutations". As before, mutations that were part of the catalytic mechanism or active site cavity were not considered further to prevent disruption to the TPS function. To narrow down the number of mutations, any mutations at similar positions in both SSCG_R6 and SdS were inserted into the FoldUnfold server to determine if they increased the packing density. Mutations were taken from all four subunits of the SdS tetramer. The mutations selected are shown in Table 4.7. All mutations were predicted to increase thermostability according to the Rosetta energy function.

Table 4.7: Packing densities, calculated by FoldUnfold, and the free energy difference between the WT and the mutant, calculated by the Rosetta energy function. The packing density values for SSCG_03688 and RoseRS_3509 were included for comparison.

Terpene Synthase	**Packing Density**	**ΔΔG (kcal/mol)**	**Terpene Synthase**	**Packing Density**	**ΔΔG (kcal/mol)**
SSCG_03688	20.785	-	RoseRS_3509	21.320	-
SSCG_R6	20.960	0.000	SdS	21.119	0.000
SSCG_R6_G46H	20.973	-2.181	SdS_G45K	21.121	-1.065
SSCG_R6_H106L	20.970	-1.439	SdS_I106L	21.118	-1.490
SSCG_R6_T207H	20.966	-1.406	SdS_R206H	21.121	-1.001
SSCG_R6_V274M	20.963	-1.128	SdS_T273K	21.113	-1.257
SSCG_R6_T311L	20.976	-1.013	SdS_V309P	21.101	-1.416

The packing densities of SSCG_03688 and RoseRS_3509 are also shown in Table 4.7 to compare with SSCG_R6 and SdS. Removal of the C-terminus to form SSCG_R6 increases the packing density for SSCG_03688 according to FoldUnfold whereas SdS has a density closer to that of the thermostable τ-muurolol synthase, RoseRS_3509. The chosen SSCG_R6 mutations all were predicted to be denser than SSCG_R6 while only G45K and R206H were predicted to have a slight increase in packing density for SdS. This might reflect the fact that only the monomer was used as an input as FoldUnfold does not accept sequences of oligomers. The position of each mutation is shown on the structures of SSCG_R6 and SdS in Figure 4.13.

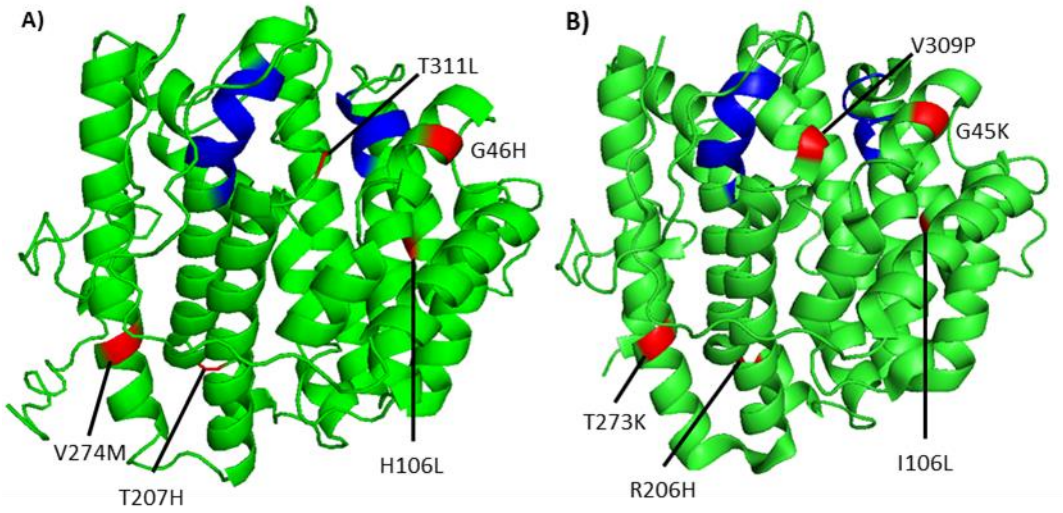


Figure 4.13: Positions of the Rosetta predicted mutations for A) SSCG_R6 and B) SdS. The Asp-rich and NSE triad motifs are shown in blue. The NSE triad is shown on the left and the Asp-rich motif on the right for each structure.

Each point mutation was introduced by overlapping PCR, the mutated genes cloned into pET28a between NheI and EcoRI sites by using restriction digestion and ligation and the resulting plasmids subsequently transformed into *E. coli* Bio*Blue* and *E. coli* BL21(DE3). Expression conditions for *sds* were as described in Baer *et al.* (2014) while the same conditions for *sscg_03688* from Hu *et al.* (2011) were used for the *sscg_R6* mutants. Large quantities of soluble protein were obtained and purified for SdS and its mutants while only the SSCG_R6_V274M and T311L mutants produced similar quantities to SdS. Expression and purification of H106L and T207H resulted in large amounts of aggregation when they were buffer exchanged and concentrated resulting in very low quantities of protein being recovered. SDS-PAGE gels of the

purification of SSCG_R6 mutants and SdS mutants are shown in Appendix Figures 13 and 14.

To determine whether the mutations affected the cyclisation mechanism, each mutant enzyme was tested by incubation at 30 °C for 30 mins with FPP. All of the SdS mutants produced selina-3,7(11)-diene as the dominant product as expected however, V309P, appeared to produce lower quantities than the rest. As for the SSCG_R6 mutants, all had similar product profiles to SSCG_R6 except T207H and H106L. T207H gave slightly lower levels of τ-muurolol than the WT while H106L produced no product at all. Work with the latter was, therefore, discontinued.

As the product profiles of most of the mutants had not changed compared to SdS and SSCG_R6, kinetic experiments were carried out using the malachite green assay (Table 4.8). For all of the SSCG_R6 mutants, the K_M values were all relatively similar to one another, other than T311L which was slightly higher. However, T311L also had a turnover number twice as fast as the other three mutants. This resulted in it having a slightly higher catalytic efficiency than the other mutants. As the kinetic parameters of SSCG_R6 were calculated using the GC-FID assay, only the catalytic efficiency can be directly compared to the mutants (Vardakou *et al.* 2014). Overall, the catalytic efficiencies of the mutants did not deviate significantly from SSCG_R6 suggesting that these mutations do not affect the kinetics.

The mutants of SdS however, showed more variation in the kinetic constants. While all of the mutants had similar K_M values to SdS, I106L had a K_M much smaller than T273K or V309P. I106L and T273K also had a k_{cat} value much lower than SdS and the other mutants. All of the resulting catalytic efficiencies were similar to the WT. This suggests, as with the SSCG_R6 data, that these mutations do not significantly effect the kinetics of SdS when comparing the catalytic efficiencies, unlike some of other studies of engineered TPS thermostability (Diaz *et al.* 2011).

Table 4.8: Kinetic parameters for SSCG_R6, SdS and the thermostability mutants using the malachite green assay. FPP concentrations used were between 0-115.4 μM except SdS_G45K which didn't have a 1.5 μM sample and SSCG_R6_T207H lacked an 11.6 μM sample due to being outliers. * The SSCG_R6 kinetic constants were calculated using the GC-FID assay.

Terpene Synthase	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($x10^{-3}$ $s^{-1}mM^{-1}$)
SSCG_R6*	16.3 ± 4.0	0.083 ± 0.006	5.1 ± 1.3
SSCG_R6_G46H	5.1 ± 0.8	0.030 ± 0.001	5.9 ± 1.0
SSCG_R6_T207H	5.3 ± 1.1	0.020 ± 0.001	3.9 ± 0.8
SSCG_R6_V274M	5.4 ± 0.6	0.029 ± 0.001	5.4 ± 0.7
SSCG_R6_T311L	8.9 ± 1.5	0.057 ± 0.002	6.4 ± 1.1
SdS	10.4 ± 2.5	0.115 ± 0.006	11.1 ± 2.8
SdS_G45K	8.4 ± 2.1	0.125 ± 0.007	14.8 ± 3.9
SdS_I106L	6.8 ± 1.1	0.073 ± 0.002	10.7 ± 1.8
SdS_R206H	8.3 ± 1.5	0.117 ± 0.004	14.0 ± 2.5
SdS_T273K	10.1 ± 1.7	0.097 ± 0.003	9.6 ± 1.6
SdS_V309P	10.6 ± 1.8	0.111 ± 0.005	10.4 ± 1.8

As there was little change in the product profiles or kinetics of the thermostability mutants, the secondary structure was assessed. Using CD, the secondary structures of each TPS and its active mutants were analysed for the expected α-helix dominant structure by wavelength scans (Figure 4.14) between 195-300 nm. All of the enzymes tested had similar profiles with minima at 208 nm and 222 nm typical of structures that are mostly α-helical. As SSCG_R6 had already been shown to have an α-helical structure (Figure 4.14A), this suggest that these mutants contain a similar secondary structure. Similarly, the crystal structure of SdS reveals that it has a typical TPS α-helical secondary structure (Baer, Rabe, Fischer, *et al.* 2014), so it is likely that the thermostability mutants also contain only α-helices and no β-sheets.

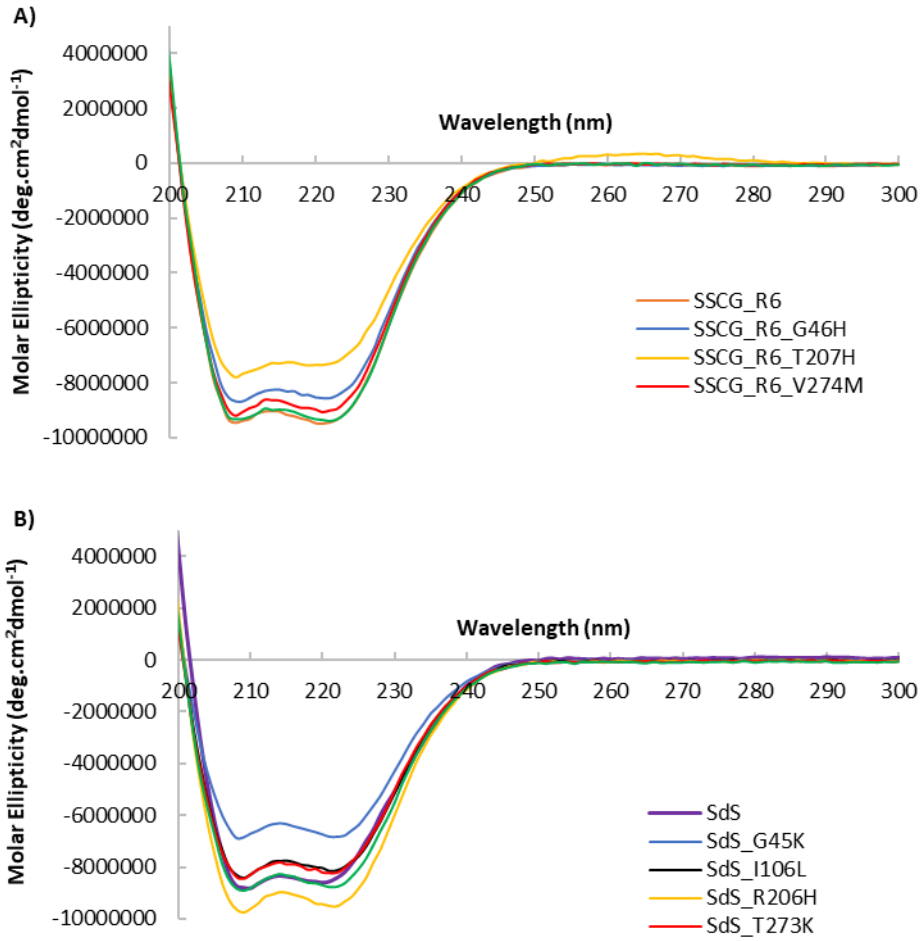


Figure 4.14: Wavelength scans of A) SSCG_R6 and B) SdS and the thermostability mutants to verify their α-helical structures using CD. The change in molar ellipticity of each TPS was measured at wavelengths between 200-300 nm at 20 °C with scan minima at 208 nm and 222 nm. SSCG_R6_H106L was not tested as not enough protein could be isolated and no activity could be seen in the GC-MS.

As all tested mutants still maintained the core α-helical secondary structure, thermal melts were performed using CD (Figure 4.15). The temperature ranges were between 20-75 °C for SdS, SSCG_R6 and all of the mutants. All of the SSCG_R6 mutants had very similar T_m values to SSCG_R6 except T207H for which the T_m value was reduced by 6.4 ^oC (Figure 4.16A). This suggests that this substitution is strongly destabilising which may relate to the aggregation observed when concentrating the protein.

Both Rosetta and FoldUnfold predicted that these mutations would increase the stability of SSCG_R6, however the experimental T_m values showed that none of these mutations produced a significant increase in the thermostability. However, SSCG_R6 was found to have a T_m of 45.8 °C which is higher than the value quoted in this thesis previously and in Styles *et al.* (2017). This was probably due to the use of fresher CD

buffer so care must be taken to only compare T_m values between the TPSs analysed in the same buffer.

Baer *et al.* (2014) calculated the melt temperature for SdS as 47.7 °C using thermal shift assays whereas here it was calculated as 56.7 °C using CD thermal melts. The SdS mutants showed variation in the melt curve shape as well as the T_m values (Figures 4.15 and 4.16B). I106L and R206H both had the biggest difference in T_m compared to the WT with decreases of >3 ^oC. V309P had a similar curve as the WT up to ~52 °C when it started unfolding at a higher rate as the temperature increased further resulting in a drop in T_m of ~1 ^oC. On the other hand, G45K and T273K showed increases in T_m of 1.1 °C and 1.2 °C, respectively. While G45K was predicted by FoldUnfold to have an increase in packing density, T273K was estimated to have a reduced density. As these programs are clearly only partially successful in predicting residue substitutions that give rise to an increase in T_m, a detailed understanding of how these mutations increase or decrease the thermostability of this TPS requires comparison of crystal structures of the WT and mutants.

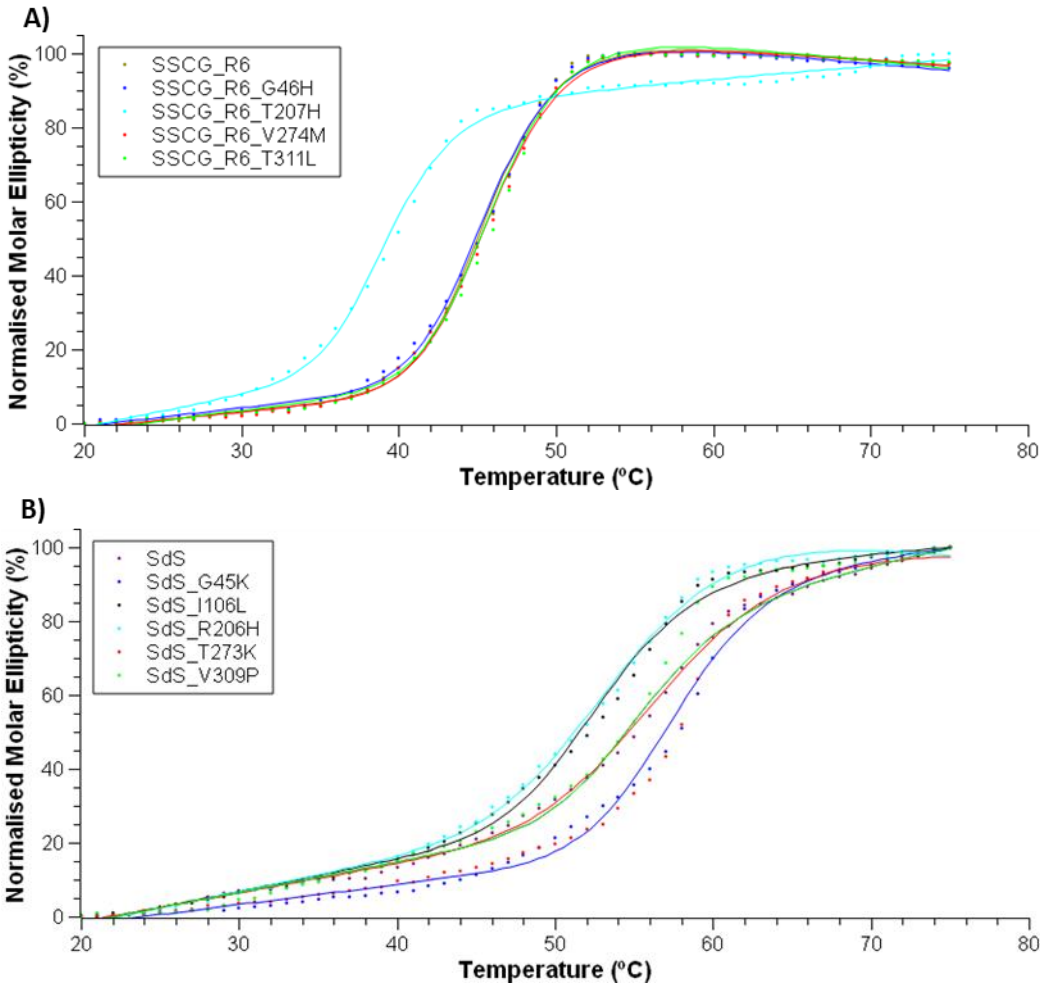


Figure 4.15: CD thermal melts of A) SSCG_R6, B) SdS and their thermostability mutants to determine each proteins T_m value. Thermal melts were performed to follow the unfolding of the secondary structure at 222 nm between 20-75 °C. The R^2 values were 0.999 for SSCG_R6 and its mutants and 0.998 for SdS and its mutants except T273K which was 0.999.

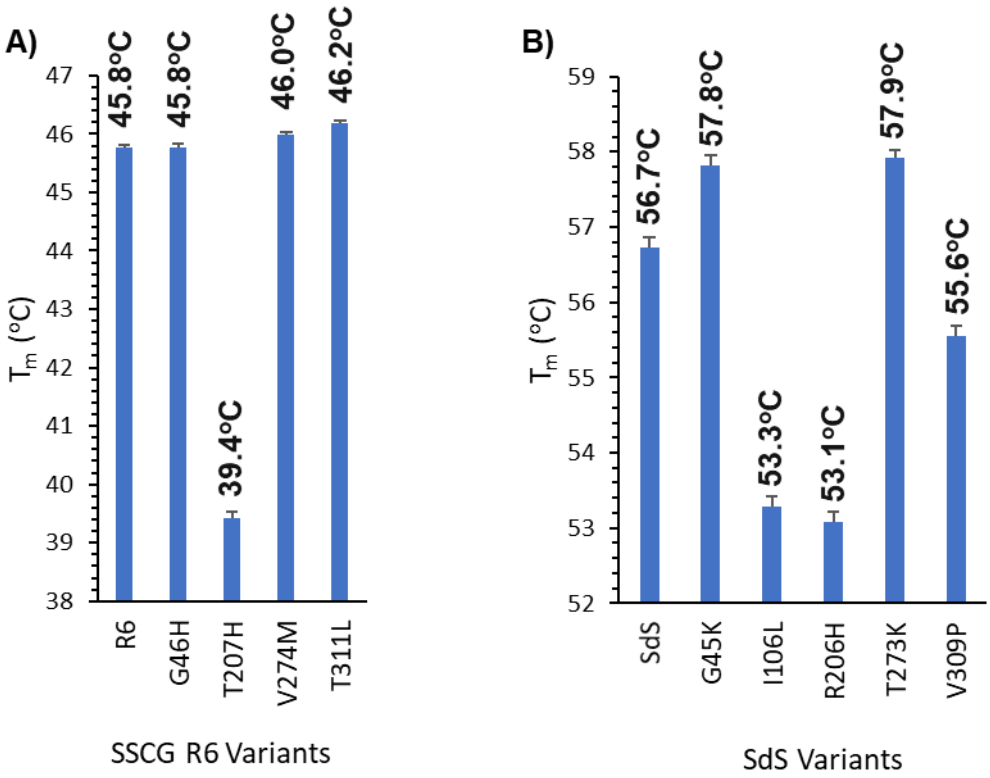


Figure 4.16: T_m values for A) SSCG_R6, B) SdS and the thermostability mutants calculated from the CD thermal melt curves in Figure 4.15. The error bars represent the deviation from the curve made by Equation 4 (Chapter 2.6.5) to the data.

Unfortunately, whilst irreversible denaturation experiments on SdS and the more thermostable mutants were trialled, the results were unreliable and not reproducible. This was not optimised further due to time constraints. Irreversible denaturation was not performed on any of the SSCG_R6 mutants as they only had very small differences in melting temperature compared with SSCG_R6.

4.3 Conclusion

Previous attempts to increase the thermostability of TPSs have shown mixed results. The T_m of BcBOT2 was increased by ~12 °C through three mutations introduced by two rounds of directed evolution using the purpald assay (Lauchli *et al.* 2013). The initial use of rational engineering to increase the thermostability of 5-*epi*-aristolochene synthase resulted in a mutant that had a T_m 45 °C higher than the WT. This mutant, however, required refolding as it was expressed in inclusion bodies and activity was

seriously compromised (Diaz *et al.* 2011). More recently, an extensive thermostability study was performed on 5-*epi*-aristolochene synthase which focussed on nine residues close to the active site that could alter the product profile of this enzyme. Incorporating multiple mutations at each of these residues, 512 mutants were screened for a change in the unfolding T_m. As well as finding mutations that increased, by up to 11 ^oC, or decreased the T_m, they determined how combinations of mutations interacted to change the T_m and the which positions individually were more likely to have a positive or negative affect on the T_m or solubility. As these mutations formed or were located near the active site cavity, many of these mutations led to an altered product profile, however, the majority of mutants were still catalytically active (O'Maille *et al.* 2008, Nartey *et al.* 2017). This chapter focussed on increasing the thermostability of a mesophilic τ-muurolol synthase, SSCG_03688, by comparison with the thermostable τ-muurolol synthases, RoseRS_3509 and Rcas_0662, as well as using computational methods to predict thermostabilising mutations away from the active site.

This chapter started by comparing the tertiary structures of the two naturally thermostable τ-muurolol synthases to a mesophilic τ-muurolol synthase (Styles *et al.* 2017). By comparing the primary and tertiary structures, it was noted that the mesostable variant had a C-terminal extension with no clear function that extended 92 residues further than the thermostable counterparts. To determine if the removal of this extension affected activity or thermostability, it was sequentially truncated in nine mutants. The C-terminal was truncated by 57 residues before stability and solubility were affected. The disruption of the only predicted secondary structure in the C-terminal, an α-helix, resulted in insoluble protein for any further truncations. The remaining soluble mutants showed that the C-terminal did not affect catalytic activity of the enzyme but did increase the T_m of the protein. The most truncated mutant, SSCG_R6, increased the T_m by ~1.2 °C and also lowered the rate of irreversible inactivation compared to the WT at 35 ^oC. This suggests that this could be a strategy for increasing the T_m of proteins, especially TPSs. To identify single mutations to further increase the thermostability of SSCG_R6, the I-TASSER model was inserted into the online program HoTMuSiC. Of the mutations predicted, seven mutations were selected away from the active site for characterisation. Of these mutations, three were soluble. Compared to SSCG_R6, G209A had a similar T_m to SSCG_R6 while G24I increased to T_m by ~1.4 ^oC. In terms of inactivation rate, G24I had lowest irreversible inactivation rate compared to both the WT and SSCG_R6. In a second round of computational prediction of mutations, the I-TASSER model of SSCG_R6 and crystal

structure of SdS were put through the Rosetta pMut script to identify thermostabilising point mutations. Five mutations were selected in similar positions in the two enzymes. Most of the mutants had similar or lower T_m values than the WT however, two mutants of SdS, G45K and T273K showed an increase in the T_m of >1 ^oC.

This work suggests that point mutations are able to increase the T_m of TPS but more mutants are required to be screened to identify mutants that are able to increase the thermostability by a larger extent. This could be achieved by performing rounds of directed evolution first to determine positions of residues which can increase the thermostability of a TPS and then make mutant combinations to determine if there is any positive synergy. The success of Lauchli *et al.* (2013) and Nartey *et al.* (2017) suggests that screening large numbers of mutants rather than using computational prediction software is more likely to find thermostabilising mutations that minimally affect catalytic activity. In future, to increase the thermostability of SSCG_03688, SSCG_R6 or SdS, a directed evolution approach should be pursued. This way many mutants could be screened and a strategy for thermostabilisation developed. This would enable more TPSs to be inserted into the *Parageobacillus* system to produce more industrially relevant terpenes at high temperatures. As SdS has a higher natural T_m value, it might be an ideal initial target. TPSs have evolved to be able to change their product distributions, the ability to predictably thermostabilise using computational methods without affecting catalytic activity is still not possible. The best route remains to use directed evolution to identify thermostable mutants of TPSs. These mutants may then inform strategies for increasing the thermostability of this group.

5 Altering the product distribution of RoseRS_3509 towards β-farnesene production

5.1 Introduction

The simplest sesquiterpene is the linear, branched hydrocarbon β-farnesene (Figure 5.1). In nature, β-farnesene is a secreted aphid alarm pheromone. It is released by aphids when under threat from a predator or parasitoid (Bowers *et al.* 1972, Francis *et al.* 2005, Beale *et al.* 2006). A range of plant species also release β-farnesene including *A. annua*, *Pinus sylvestris* and *Mentha asiatica* (Picaud *et al.* 2005, Köpke *et al.* 2010, Yu *et al.* 2013). In addition to repelling the aphids, many aphid predators have adapted to recognise the release of β-farnesene as an indicator for the presence of their prey (Francis *et al.* 2004, Harmel *et al.* 2007, Zhou *et al.* 2016).

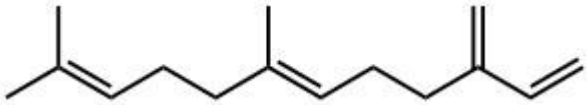


Figure 5.1: Structure of β-farnesene.

In industry, β-farnesene has many applications; it is used as a feedstock for producing squalene for cosmetics, rubber and other polymers, or even as a biofuel (McPhee *et al.* 2014, Schofer *et al.* 2014, Leavell *et al.* 2016). It is also used for solvents, surfactants, emulsifiers and vitamin precursors (https://farnesene.net/). Chemical synthesis of β-farnesene is unsustainable due to the chemicals involved in catalysis (Mandai *et al.* 1983, Poppe *et al.* 1987) and there are no examples of commercial β-farnesene extraction from plants probably reflecting the low yields of terpenes extracted from plants (Peralta-yahya *et al.* 2011, Mewalal *et al.* 2017). Several commercial recombinant microbial processes for terpene production exist but engineered *S. cerevisiae* containing a heterologous mevalonate pathway is able to produce the highest titres. This system, developed by Amyris, has been optimised for the production of β-farnesene under its commercial name, Biofene®, with the most recent quoted yield being 130 g/L of yeast culture (Meadows *et al.* 2016). This system is also used for producing other useful terpene compounds including the anti-malarial, artemisinin, and the artificial sweetener, RebM (https://amyris.com/). Titres of

terpenes from engineered native and heterologous *E. coli* systems have shown promise but have not reached commercialisation yet (Ajikumar *et al.* 2010, Wang, Park, *et al.* 2016, Li *et al.* 2017, Zada *et al.* 2018). The highest recorded yield of β-farnesene from *E. coli* to date was 8.74 g/L using a heterologous mevalonate pathway with increased expression of isopentyl diphosphate isomerase and farnesyl diphosphate synthase and using glycerol as a substrate (You *et al.* 2017). All of the current commercial platforms, as well those being developed, use glucose, sucrose or glycerol as carbon sources (Alonso-Gutierrez *et al.* 2013, Lv *et al.* 2016, Wang, Park, *et al.* 2016, Kim *et al.* 2017, Zada *et al.* 2018); the Amyris platform uses sucrose from sugarcane (McPhee *et al.* 2014). An alternative to using a carbohydrate that might also be used as a food stuff would be to use waste lignocellulosic biomass from the agricultural industry (Cripps *et al.* 2009, Tan *et al.* 2016), such as corn stover, bagasse or straw. This would require different microbial hosts capable of growth on biomass hydrolysates, such as *Parageobacillus*/*Geobacillus* species.

For β-farnesene production in *S. cerevisiae*, Amyris use an over-expressed *A. annua* β-farnesene synthase (Chandran *et al.* 2011). If β-farnesene was to be produced at high temperatures, one would need to engineer a thermostable β-farnesene synthase from an already thermostable TPS as no naturally thermostable β-farnesene synthases have been characterised. As discussed in chapter 1.4, some of the advantages of terpene production at high temperatures include higher substrate solubility, lower contamination risk and lower cooling costs. To change an enzyme's terpene product, information about the enzyme's structure and mechanism needs to be understood. With the diphosphate of FPP bound by the conserved motifs at the top of the active site, the substrate's prenyl chain is forced into a substrate-like conformation in the shallow active site. On removal of the diphosphate and formation of the carbocation, this cation is chaperoned through the appropriate cyclisation steps by the residues (Mainly aromatic) of the active site to form the desired product. Many TPSs have a high specificity for generating a single product with particular regio- and stereochemistry, typically with only one or two minor products. However, some show lower specificities and higher promiscuities, like γ-humulene synthase which produces 52 different sesquiterpenes (Steele *et al.* 1998). The specificity depends on the active site and its ability to stabilize and mobilize the carbocation through specific rearrangement steps. If the active site is not specific, the substrate can bind in different conformations and alternative carbocations form resulting in products with different structures and permitting promiscuity. In the case of γ-humulene synthase (Figure 5.2A), as well as producing the monocyclic γ-humulene, the enzyme also

produces a complex mixture of linear, bicyclic and tricyclic sesquiterpenes suggesting multiple different cyclisation reactions are occurring. In many cases, the other terpenes are stabilized intermediates caused by premature termination of different carbocation intermediates. This is the case for *epi*-isozizaene synthase which produces five minor terpenes as a result of deprotonation of various carbocations on the way to *epi*-isozizaene (Lin and Cane 2009, Aaron *et al.* 2010). Specific functions in this enzyme group are believed to have evolved from more promiscuous enzymes through divergent evolution. The changes made to increase specificity are thought to rely on only a few mutations in residues known as plasticity residues. These residues dictate the shape of the active site cavity so they either line the active site or are neighbouring residues (Figure 5.2B, Yoshikuni, Ferrin and Keasling, 2006).

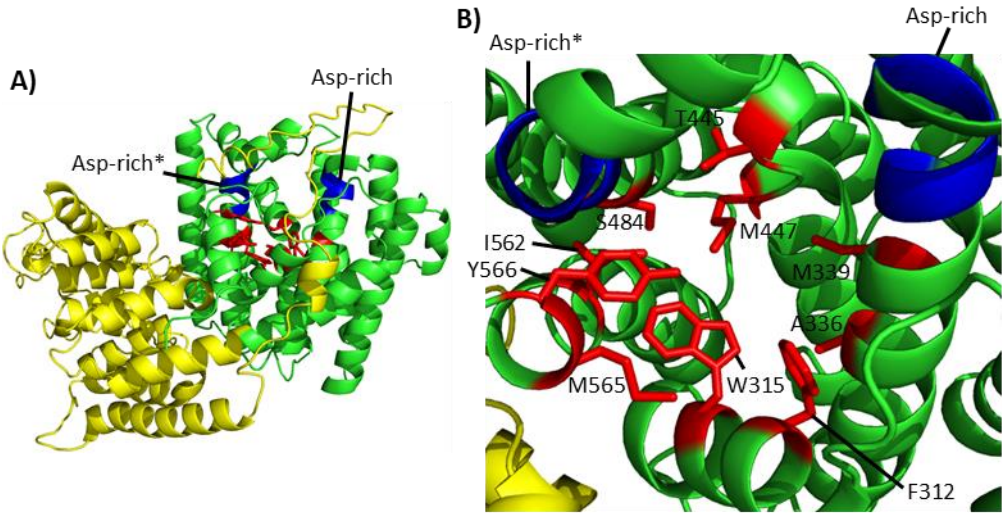


Figure 5.2: Homology model of A) γ-humulene synthase and B) its active site. A) The homology model of γ-humulene synthase made using I-TASSER with the β- and α-domains shown in yellow and green, respectively. The Asp-rich motifs are shown in blue. A rare second Asp-rich* motif replaces the NSE triad. B) The γ-humulene synthase active site looking down from the top of the active site. The red residues are those labelled as plasticity residues by Yoshikuni *et al. (2006)* as they alter the product specificity of this TPS. Not all of the active site residues are plasticity residues and have not been labelled in this figure.

Multiple studies have shown that only a few mutations are required to alter the specificity of a TPS towards an alternative terpene product. Yoshikuni *et al.* (2006) took the promiscuous γ-humulene synthase and performed site saturation mutagenesis on the 19 residues making up the active site cavity. They identified that 1-5 mutations were required to switch the specificity from γ-humulene to one of its

minor products such as β-farnesene and sibirene. They also determined mutations that increased the specificity towards γ-humulene from 45% to 86% of the total product compared to the WT. Altered product distribution was also engineered into δ-cadinene synthase, which predominantly produces δ-cadinene with minute amounts of germacrene D-4-ol (Chen *et al.* 1995, Yoshikuni, Martin, *et al.* 2006). Germacrene D-4-ol is a stabilized version of the monocyclic cation used in δ-cadinene synthesis. It was determined that a single mutation, N403P or L405H, enabled germacrene D-4-ol to become the dominant product and both mutations combined increased the solubility of the new TPS (Yoshikuni, Martin, *et al.* 2006). However, mutations around the active site have mixed results in terms of activity of the resultant enzyme. In most cases, the k_{cat} decreases while the K_M tends to increase causing an overall decrease in k_{cat}/K_M. By mutating the active site, the TPS loses specificity as each residue is in a particular position to enable the mechanism to proceed. Changing the active site contour not only changes the binding of the substrate but also the ability to cyclise the substrate (Deligeorgopoulou and Allemann 2003, Yoshikuni, Martin, *et al.* 2006, Aaron *et al.* 2010, Li *et al.* 2014, Blank *et al.* 2017).

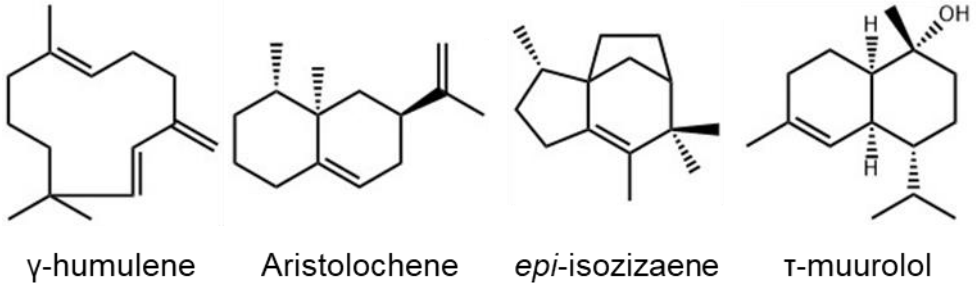


Figure 5.3: Structures of γ-humulene, aristolochene, *epi*-isozizaene and τ-muurolol. The initial cyclisation mechanisms of NPP are 1,11-, 1,6- and 1,10-cyclisation for γ-humulene, *epi*-isozizaene and τ-muurolol, respectively. The initial cyclisation mechanism for aristolochene is a 1,10-cyclisation of FPP.

Several studies have enabled TPSs to become β-farnesene synthases through a single mutation. In all of the studies, the mutation of an aromatic residue, around the base of the active site enabled the production of β-farnesene. In γ-humulene, aristolochene and *epi*-isozizaene synthases (Figure 5.3), these mutations were W315P, Y92A or F96A, respectively; a consistent pattern despite having plant, fungal and bacterial origins (Deligeorgopoulou and Allemann 2003, Yoshikuni, Ferrin, *et al.* 2006, Aaron *et al.* 2010). These mutations allowed the reaction to be immediately terminated by deprotonation after diphosphate release by binding the substrate in a

more open conformation. However, only Deligeorgopoulou and Allemann (2003) investigated how the mechanism may have changed because of the mutation. This study on aristolochene synthase from *Penicillium roqueforti* hypothesized that the removal of the tyrosine residue at the base of the active site increased the depth of the active site cavity. This enabled the once folded substrate to bind in an extended conformation preventing any cyclisation reactions as the reactive carbons are no longer in proximity to one another. This results in deprotonation to mainly β-farnesene and some α-farnesene. As β-farnesene synthases have been made through this common mechanism of a single mutation, we hypothesised that a similar mutation could be made to create the first thermostable β-farnesene synthase.

To increase the number of terpenes produced at high temperatures, this work aimed to engineer RoseRS_3509 to produce β-farnesene as its dominant product (Styles *et al.* 2017). This enzyme has a relatively high k_{cat} for a bacterial TPS so the expected reduction in catalytic activity should still yield a competent enzyme. It also aimed to determine which residues in the active site are important for τ-muurolol production. The conserved motifs, the Asp-rich motif, pyrophosphate sensor, NSE triad and RY motif mentioned in chapter 1.3, were not mutated due to this causing a substantial loss in activity observed when this was done in other enzymes (Seemann *et al.* 2002, Lin and Cane 2009, Baer, Rabe, Fischer, *et al.* 2014). Also, the less well characterised pyrophosphate sensor, R175, and reaction initiator, G179, are also vital and were therefore not mutated (Baer, Rabe, Fischer, *et al.* 2014). To achieve this, attempts were made to generate structural information through crystal structures and models, to determine which residues to target. Residues, especially aromatics, in the base of the active site were mutated and the mutant's product distribution determined. Any new thermostable TPSs were then inserted into the *Parageobacillus* platform to determine whether a new terpene could be produced at high temperatures.

5.2 Results and Discussion

5.2.1 Thermostable τ-muurolol synthase crystallography

Since the first bacterial TPS, pentalenene synthase, was crystallised in 1997, only a further seven unique crystal structures have been solved (Lesburg *et al.* 1997, Aaron *et al.* 2010, Köksal *et al.* 2012, Baer, Rabe, Citron, *et al.* 2014, Baer, Rabe, Fischer, *et al.* 2014, Janke *et al.* 2014, Harris *et al.* 2015, Karuppiah *et al.* 2017). This is small compared to the large number of new TPS sequences shown to exist in bacteria (Chapter 4) and from extensive genome mining (Komatsu *et al.* 2008, Yamada *et al.* 2012, 2015). As Rcas_0662 and RoseRS_3509 were the first TPSs to be characterised as thermostable (Styles *et al.* 2017), their structures would be useful for understanding their thermoactivity. This would include information about the catalytic mechanism, substrate binding and what makes it thermostable.

To produce sufficient protein for crystallisation, *E. coli* BL21(DE3) containing pET28a constructs of *rcas_0662* and *roseRS_3509* were expressed in 6L of LB culture. The pET28a plasmid added an N-terminal His-tag to Rcas_0662 and RoseRS_3509. To extract the proteins, the cells were sonicated before removal of the cell debris by centrifugation. Extracted proteins were initially purified by affinity chromatography using cobalt affinity resin followed by size exclusion chromatography. After affinity chromatography, there were still multiple SDS-PAGE bands evident, other than those corresponding to Rcas_0662 and RoseRS_3509, many of which were removed through size exclusion chromatography (Figure 5.4). However, even after this, two additional bands consistently appeared on the SDS-PAGE gels other than the target polypeptide. The larger band (~80.0 kDa) was believed to be an oligomer of Rcas_0662 while the smaller bands (~23.0 kDa and ~24.0 kDa) were believed to be a protein degradation product of the targets or an *E. coli* protein that associated with them. The smaller band was not removed by concentration in Amicon 30 kDa MWCO Ultra-15 centrifugal filter units. Rcas_0662 was concentrated to 8.7 mg/ml and RoseRS_3509 to 13.3 mg/ml.

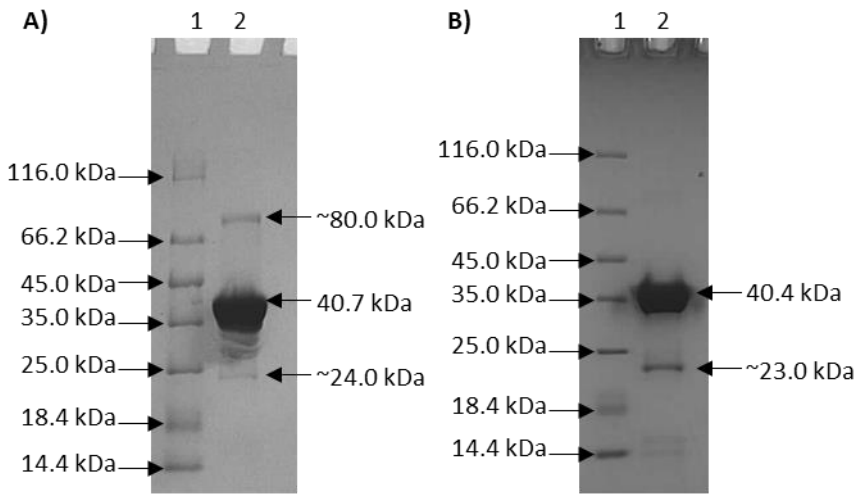


Figure 5.4: SDS-PAGE gels of A) Rcas_0622 (40.7 kDa) and B) RoseRS_3509 (40.4 kDa) purified by affinity chromatography and size exclusion chromatography. Lane 2 for each image was the protein sample and the concentration of protein added to the gels were 1.0 mg/ml for Rcas_0662 and 1.3 mg/ml for RoseRS_3509. Lane 1 for both images was the unstained protein molecular weight marker.

As both protein preparations were fairly pure, the pre-crystallisation test (PCT, Hampton Research) was used to determine whether crystallisation should be attempted. Varying concentrations between 2.0-8.7 mg/ml were tested with only Rcas_0662 being suggested for crystallisation at 2.2 mg/ml. RoseRS_3509 produced heavy precipitation under all conditions and protein concentrations tested. Sitting-drop crystallisation screens of Rcas_0662 were carried out using JCSG-plus™, SG1™, Clear Strategy™ Screen I, Clear Strategy™ Screen II, Structure Screen I & II, PACT premier™, MIDAS™ and ProPlex™ (Molecular Dimensions Ltd). The Rcas_0662 concentration was calculated before adding to the crystallisation plates as 2.7 mg/ml. Plates were stored at 16 ^{o}C. Over several months, droplets were either clear or had varying amounts of precipitation but no crystals were identified.

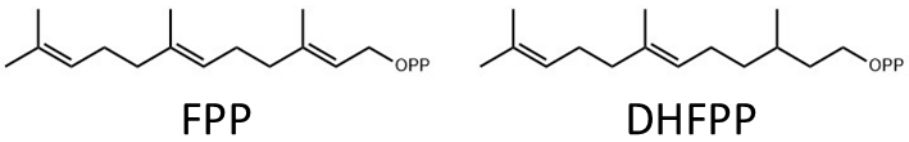


Figure 5.5: Structures of FPP and 2,3-dihydrofarnesyl diphosphate (DHFPP). The 2,3-carbon double bond stabilises the initial carbocation formed when the diphosphate group leaves. As this double bond is not present in DHFPP, it is unreactive as the diphosphate group cannot leave and will bind to a sesquiterpene synthase active site.

To try to encourage crystal formation by stabilising RoseRS_3509, the substrate analogue of FPP, 2,3-dihydrofarnesyl diphosphate (DHFPP, Figure 5.5) was kindly synthesised by Robert Chapman (Department of Chemistry, University of Bath, UK). This analogue was used to crystallise selinadiene synthase (SdS) as well as determine the residues associated with the initial ionisation and overall cyclisation mechanisms (Baer, Rabe, Fischer, *et al.* 2014). As advised by Philipp Baer (Personal communication), RoseRS_3509 was diluted to 10 mg/ml and added to an aliquot of freeze-dried DHFPP to create a final concentration of 1 mg/ml of DHFPP. This solution was incubated at 4 °C for 1 hr before pelleting out all of the precipitate. The remaining RoseRS_3509 was quantified at 6 mg/ml and PCT suggested that it could be used for crystallisation. The crystallisation plates used were Morpheus®, PACT premier™, SG1™ and LMB Crystallization Screen™ with sitting drops set up at 2:1, 1:1 and 1:2 ratios of protein to precipitant and using a 50 µl precipitant reservoir. These plates were kept at 16 °C over five months, however, no protein crystals formed.

5.2.2 Mutant determination

As crystals of RoseRS_3509 were not obtained, the structure was modelled instead. Initially to determine which residues to mutate, the active site of RoseRS_3509 was modelled using I-TASSER (Figure 5.6). As expected for binding the prenyl chain of the substrate, the lower part of the active site was shown to contain mainly hydrophobic residues like valine, isoleucine and alanine. This included several aromatic residues that are highly likely to stabilize and manipulate the intermediate carbocations during backbone rearrangement. Several of these aromatic residues lie at the base of the active site. The upper section mainly contained charged and polar residues for coordinating the Mg^{2+} ions and the diphosphate moiety.

As the probable cyclisation mechanism for τ-muurolol generation is already known (Figure 5.7), docking of both FPP and DHFPP into the I-TASSER model of RoseRS_3509 were attempted by Rory Crean (University of Bath, UK) to try and determine which residues to mutate. However, neither were able to be positioned in the active site using this model. Instead, to target production of β-farnesene rather than τ-muurolol as the product of RoseRS_3509, the primary sequence was aligned with the γ-humulene, aristolochene and *epi*-isozizaene synthases mentioned above that had been mutated to become β-farnesene synthases (Figure 5.8). For the

alignment with γ-humulene synthase, the sequence for the β-domain was removed and only the α-domain was used.

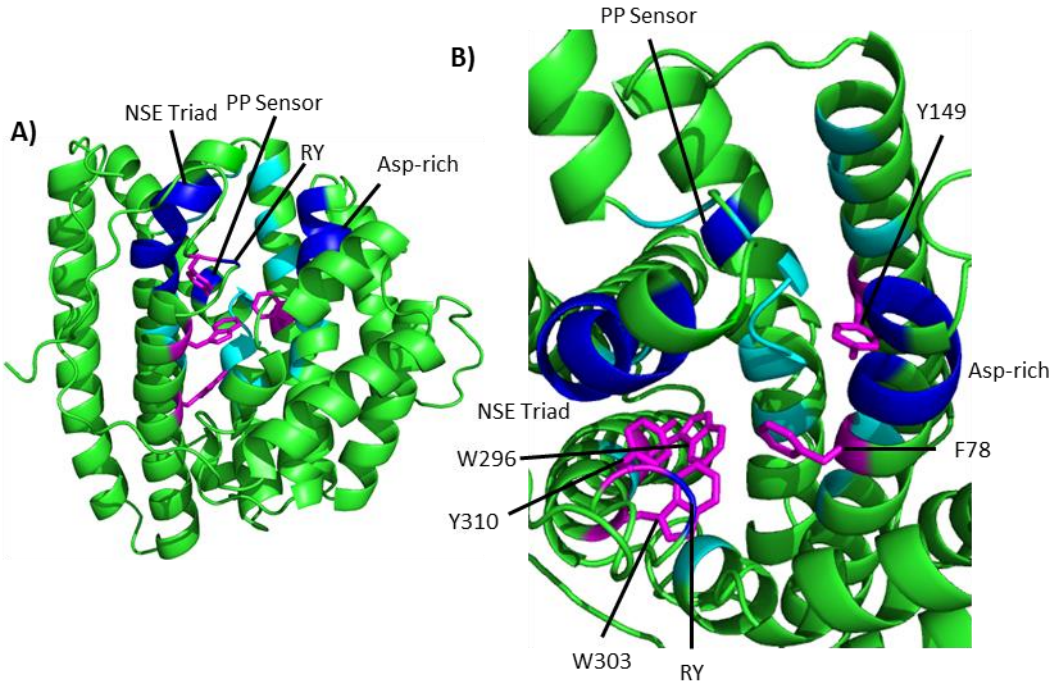


Figure 5.6: Homology model of the A) structure and B) active site of RoseRS_3509. A) The model of RoseRS_3509 created using I-TASSER with the Asp-rich, pyrophosphate (PP) sensor, NSE triad and RY motifs labelled in blue and G179 in red. B) Looking down from the top of the RoseRS_3509 active site. The α-carbon of the residues predicted to line the active site are shown in cyan and the aromatic residues, F96, Y149, W296, W303 and Y310 shown in magenta.

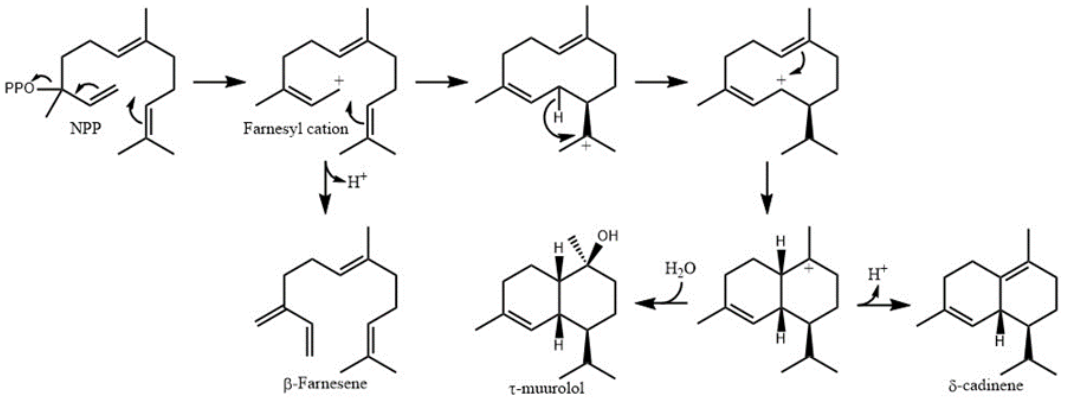


Figure 5.7: Predicted cyclisation mechanism for the products made by RoseRS_3509 and how β-farnesene can be made by deprotonation of the farnesyl carbocation. The manipulation of FPP to NPP is described previously. The carbocations are terminated either by deprotonation or hydroxylation. Image adapted from Dickschat et al *(2014)*.

```
AS             1   MATSTET-----ISSLAQPFVHLENPINSPLVKETIRPRNDTTITPPPTQ
EIZS           1   MVHAFPHGTTATPTAIAVP-----PSLRLPVI---------------EAA
GMS            240 VQRWE-------ARSFIEIFGQIDSELKSNLSKKMLELA--------KLD
RoseRS_3509    1   MDRVD-------GAQIVYPFTGAI--------------------------
                       :            :

AS             46  WSYLCHPRVKEVQDEVDGYFLENWKFPSFKAVRTFLDAKFSEVTCLYFPL
EIZS           31  FPRQLHPYWPKLQETTRTWLLEKRLMPADKVEEYADGLCYTDLMAGYYLG
GMS            276 FNILQCTHQKELQI-ISRWFADSSIASLNFYRKCYVEFYFWMAAAISEPE
RoseRS_3509    18  -----SPYAGDVDQATLIWAEASG-LLTDGLRQKSQRLQYGVLAARAYPQ
                         .   .::    :   .        .     :    .

AS             96  ALDDRIHFACRLL-TVLFLIDDVLEHMSFADG-E---AYN--NRLIPI--
EIZS           81  APDEVLQAIADYS-AWFFVWDDRHDRDIV-HG-R---AGA-WRRLRGLLH
GMS            325 FSGSRVAF--TKIAILMTMLDDLYDTHGTLDQLK---IFTEGVRRWDV--
RoseRS_3509    62  ADRDTLQIAADWI-AWLFFMDDQCDEAGIGRDPQRLAALH--ERFLAV--
                     . :         : . **  :        .        *   :

AS             137 SRGDVLPDRTKPEEFILYDLWESMRA--------HDAE-----------L
EIZS           124 TALDSPGDHLHHEDTLVAGFADSVRRLYAFLPATWNAR-----------F
GMS            368 --------------SLVEGLPDFMKIAFEFWLKTSNELIAEAVKAQGQDM
RoseRS_3509    107 LEGD-PPDSG--DWNLTRALADIRRRLAARATDDWLRR-----------F
                               :   : :  :                        :

AS             168 AN-----EVLEPTFVFMRAQTDRARLSIHELGHYLEYREKDVGKALLSAL
EIZS           163 AR-----HFHTVIEAYDREFHNRTRGIVPGVEEYLELRRLTFAHWIWTDL
GMS            404 AAYIRKNAWERYLEAYLQDAEWIATGHVPTFDEYLNNGTPNTGMCVLNLI
RoseRS_3509    143 GE-----HVRLYFTANRWEAANRCQSIVPNVATYCAARLFSGAVYACFDL
                   .           .           :  .  *       .     :

AS             213 MRFSMGLRLSAD----------ELQDMKALEANCAKQLSVVNDIYSYDKE
EIZS           208 LEPSSGCELPDA--------VRKHPAYRRAALLSQEFAAWYNDLCSLPKE
GMS            454 PLLLMGEHLPIDILEQIFLPSRFHHLIELASR-------LVDDARDFQAE
RoseRS_3509    187 IELAAGIDLPFY--------ARYHAAVQQLERTANNIICWCNDMLSYPKE
                        *  *.                .            :*  .   *

AS             253 EEASRTGHKEGAFLCSAVKVLAEESKLGIPATKRVLWSMTREWE----TV
EIZS           250 IAGDEV-HNLGISLITHHSLTLEEA------IGEVRRRVEECIT----EF
GMS            497 KDHGDL-SCIECYLKDHPESTVEDA------LNHVNGLLGNCLLEMNWKF
RoseRS_3509    230 MQHGDV-HNLVLAIRQEYQCSLSEA------LQQALLLHDRETD----TF
                      .        :    .   .::        ..     .        .

AS             299 HDEIVAEKIASPD-------------GCSEAAKAYMKGLEYQMSGNEQWS
EIZS           289 ----LAVERDALRFADELADGTVRGKELSGAVRANVGNMRNWFSSVYWFH
GMS            540 ----LKKQDSVPLS----------CKKYSF--HVLARSIQFMYNQGDGFS
RoseRS_3509    269 ----MKTQKQLPRF----------NPAVDMALERYIDGLQYWICANRDWS
                       :  :                   .   .    .:.        :

AS             336 KTTRRYN---------------------------
EIZS           335 HESGRYM------VDSWDDRSTPPYVNNEAAGEK
GMS            574 ISNKVIKDQVQKVLIV-----PVPI---------
RoseRS_3509    305 LTAMRYA------LSG-----TEPVLRTRFAH-S
```

Figure 5.8: Primary sequence alignment of RoseRS_3509, aristolochene synthase (AS), *epi*-isozizaene synthase (EIZS) and the α-domain of γ-humulene synthase (GHS). The residues labelled in red from AS (Y92A), EIZS (F96A) and GHS (W315P) are those that were mutated to create β-farnesene synthases and those residues in which they align to in RoseRS_3509. Alignments of the protein primary sequences were performed using T-coffee (EBI). The crystal structure of *epi*-isozizaene synthase is shown in Figure 1.9 in the introduction. Uniprot accession numbers: AS, Q03471; EIZS, Q9K499; GHS, O64405.

The sequence alignment showed that G52, A58 and L77 of RoseRS_3509 aligned with the β-farnesene mutations of the other enzymes (Figure 5.8). However, only L77 was predicted to line the RoseRS_3509 active site. However, this residue was not a large aromatic that could cause a major change in the active site volume through mutation. As no convincing residues were identified, the RoseRS_3509 I-TASSER model was aligned with the crystal structures of *epi*-isozizaene (3KB9) and aristolochene (1DI9) synthases and the I-TASSER model of γ-humulene synthase. From these alignments, only F96 from *epi*-isozizaene synthase aligned with an aromatic residue, F78, in RoseRS_3509 (Figure 5.9A). F78 sits in a similar position to F96 in *epi*-isozizaene synthase at the side of the base of the active site. The Y92 residue of aristolochene synthase and W315 residue of γ-humulene synthase overlaid with A58 and L54 from RoseRS_3509, respectively. Of the two, only L54 was predicted to be a part of the active site (Figures 5.9B and 5.9C).

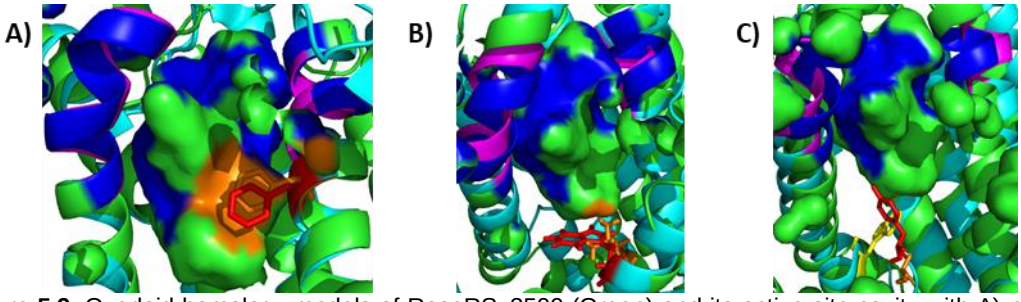


Figure 5.9: Overlaid homology models of RoseRS_3509 (Green) and its active site cavity with A) *epi*-isozizaene synthase (Cyan), B) γ-humulene synthase (Cyan) and C) aristolochene synthase (Cyan). The Asp-rich and NSE triad motifs are coloured in blue for RoseRS_3509 and magenta in the other enzymes. The orange residues are A) F78, B) L54 and C) A58 in RoseRS_3509 and the red residues are A) F96, B) W315 and C) Y92 in *epi*-isozizaene, γ-humulene and aristolochene synthases, respectively. The yellow residue in C) is W296 from RoseRS_3509 which overlaps with Y92 in aristolochene synthase. Some helices were removed to create a clearer image of the active sites.

Also, in the RoseRS_3509 model, W296 was recognised as covering the majority of the base of the active site (Figure 5.9B). Although, not aligning with the mutations of the sequences or structures of the other TPSs, it was predicted to overlap with Y92 from aristolochene synthase in its position compared to the active site (Figure 5.9C). On modelling the mutations of F78 and W296 to alanine, both models showed a large increase in the active site volume. When mutating Y92 in aristolochene synthase, Deligeorgopoulou and Allemann (2003) progressively replaced tyrosine with smaller residues from valine to cysteine to alanine. This change saw a step-wise increase in the proportion of β-farnesene in the total product from 0% in the WT to 73% in Y92A.

However, for *epi*-isozizaene synthase this was not the case. Mutation of F96 to smaller residues only resulted in β-farnesene production when switched to alanine. Mutation to valine and leucine still resulted in the cyclic products, γ-bisabolene and (+)-prezizaene, being dominant with no β-farnesene observed (Aaron *et al.* 2010, Li *et al.* 2014). To determine how these two residues from RoseRS_3509 affect function, F78 was mutated to alanine and valine and W296 to alanine and leucine. Valine was chosen for F78 as F96V in *epi*-isozizaene synthase showed a larger shift to terpene structures caused by early termination of the cyclisation than F96L (Li *et al.* 2014). As a residue in the same position as W296 had not been mutated in any other TPS, leucine was chosen as this would represent a similar change in cavity size to F78V. Also, residues which formed the lower hydrophobic part of the active site were mutated to determine how important their role was in τ-muurolol cyclisation. This included L54, I74, L77, C184, N300 and W303 which were all mutated to alanine individually. Although forming the top of the active site rather than the base, Y310 was included in the mutations as it was an aromatic. As Y310 forms part of the conserved RY motif, it was expected to lose activity as both the arginine and tyrosine are important for substrate binding (Cane *et al.* 1995, Aaron *et al.* 2010, Baer, Rabe, Fischer, *et al.* 2014). However, the mutation of both residues in the RY motif in hedycaryol synthase did not affect the activity compared to the WT (Baer, Rabe, Citron, *et al.* 2014). No residues in close proximity to R175 and G179, were mutated. As the aromatic residue, Y149, neighbours these residues it was not mutated to avoid disrupting reaction initiation. R175 is predicted to be the pyrophosphate sensor in RoseRS_3509 while the carbonyl of G179 is predicted to initiate the removal of the diphosphate to form the reactive carbocation (Baer, Rabe, Fischer, *et al.* 2014). Lastly, Y59, was included as it sits below W296 at the base of the active site. Mutation of this residue may alter the position of W296 in the active site by creating a cavity behind it. This might enable extension of the active site for altered substrate binding. All of the mutated residues are shown in Figure 5.10.

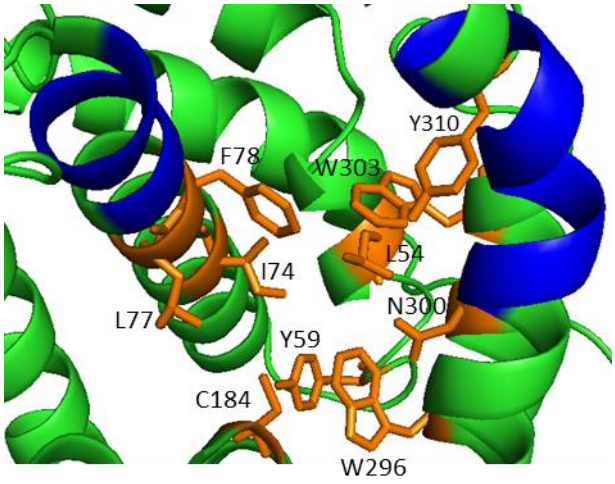


Figure 5.10: Positions of the residues targeted for mutation in the RoseRS_3509 active site. The backbone in blue is the Asp-rich (Left) and NSE triad (Right) motifs at the top of the active site. Some residues were hidden to enable a full view of the residues targeted for mutation to be visualised.

5.2.3 Mutant properties

From the *E. coli* codon optimised *roseRS_3509* gene, each mutation was introduced using overlapping PCR and cloned between the NheI and EcoRI restriction sites in pET28a. The presence of the desired mutations was confirmed by sequencing (GATC Biotech, Germany). The same optimal conditions for expression of *roseRS_3509* were used for the mutants (Styles *et al.* 2017). All mutants were expressed in the soluble fraction and the SDS-PAGE gels from the affinity purification are shown in Appendix Figure 15. Korakot Ratsomrong (Former Masters student, Leak laboratory) cloned, expressed and purified L54A, Y59A, I74A, C184A and W296L and Charlotte Tidd (Former undergraduate student, Leak laboratory) cloned, expressed and purified F78A, F78V, N300A, W303A and Y310A under my close supervision.

Wavelength scans using circular dichroism for each of the mutants suggested that the α-helical nature of a TPS was maintained after mutation of the active site as there were the characteristic minima at 208 nm and 222 nm of a dominant α-helical structure (Figure 5.11). To assess whether these active site mutations affected thermostability, thermal melts followed by circular dichroism were performed between 20-90 ^{o}C. Compared to the WT, all of the mutated proteins showed a decrease in thermostability according to the T_m values (Figure 5.12 and Table 5.1). The mutant with the lowest T_m was L77A which had an ~10 °C decrease compared to the WT. Similarly, the values of decreased van't Hoff enthalpy (ΔH) values for all of the mutants shows that they were more thermodynamically unstable than RoseRS_3509.

This suggests that mutations in the active site lower the stability of RoseRS_3509. Of the mutants, F78A and C184A were the most thermostable. The point at which the gradients of the thermal melt curves of L54A and L77A begin to increase exponentially was at a lower temperature than the WT resulting in lower T_m values overall. This suggests that these mutant proteins are slightly more unstable at higher temperatures than the other mutants. However, as no mutants showed a large decrease in the T_m, the product distributions of the WT and each mutant were analysed at 37 °C.

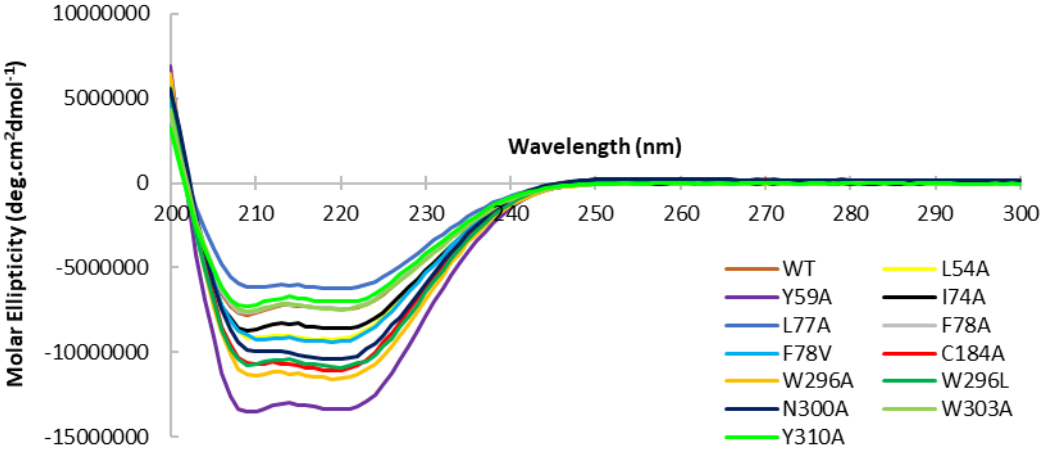


Figure 5.11: Wavelength scans of RoseRS_3509 (WT) and its mutant proteins between 200-300 nm to confirm the secondary structures using CD at 20 °C. Minima were at 208 nm and 222 nm. The concentration of each of the proteins was 1 µM.

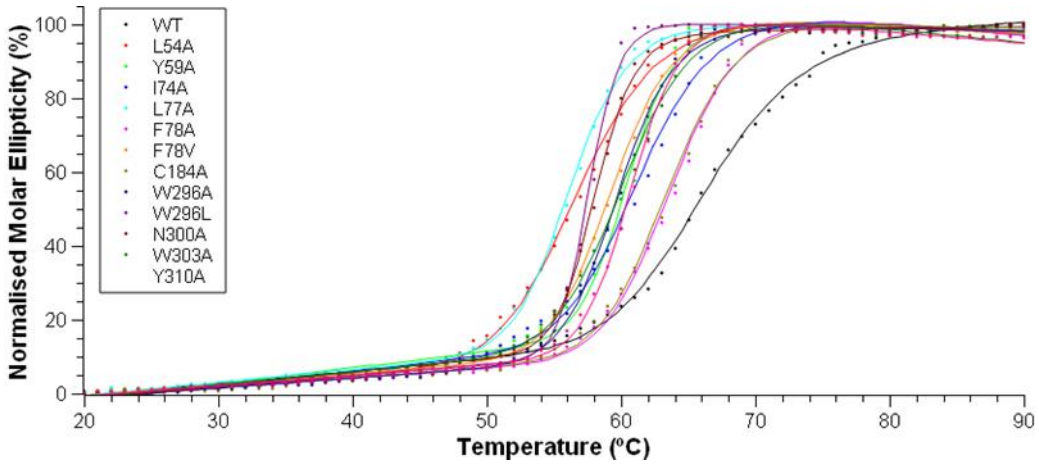


Figure 5.12: Normalised unfolding curves of RoseRS_3509 (WT) and the variant proteins between 20-90 °C at 222 nm using CD to calculate the T_m values. The R^2 value was 0.999 for all mutants except Y59A which was 0.998.

Table 5.1: Melt temperatures (T_m) and van't Hoff enthalpies (ΔH) of the RoseRS_3509 active site mutants from Figure 5.12. The errors are the errors of fitting the data to the curve created by Equation 4 (Chapter 2.6.5).

Terpene Synthase	T_m (°C)	ΔH (kJ/mol)
RoseRS_3509	66.5 ± 0.09	32.2 ± 0.7
L54A	57.1 ± 0.07	37.0 ± 0.7
Y59A	60.6 ± 0.09	55.7 ± 2.2
I74A	61.6 ± 0.09	39.2 ± 1.1
L77A	56.3 ± 0.04	50.1 ± 0.7
F78V	59.5 ± 0.03	47.5 ± 0.6
F78A	64.0 ± 0.06	45.3 ± 1.3
C184A	63.7 ± 0.07	43.8 ± 1.0
W296L	57.6 ± 0.05	93.8 ± 3.5
W296A	59.9 ± 0.05	50.4 ± 0.9
N300A	58.4 ± 0.03	73.2 ± 1.4
W303A	60.5 ± 0.06	43.6 ± 1.0
Y310A	60.9 ± 0.05	60.5 ± 1.1

Initially, RoseRS_3509 and its mutants were assayed at 37 °C for 1 hour with FPP (Figure 5.14). The extracted ion content (XIC) profiles of the GC-MS were used qualitatively to determine which terpenes were being produced (Figure 5.13 and Table 5.2). The ion fragmentation spectra were compared to the NIST database to determine the terpene structure associated with each peak. The flame ionisation detector (FID) data was used quantitatively but it was not as sensitive as the XIC and therefore not all peaks that were visualised on the XIC were seen on the FID. The NIST database was able to predict the peaks for β-farnesene, cadinene-like terpenes, germacrene D-4-ol and τ-muurolol with high confidence (Figure 5.13). Only farnesene could be confirmed by using the farnesene, mixture of isomers sample (Sigma-Aldrich, UK, Appendix Figure 8) as no authentic standards were available of the other terpene compounds although the retention time of τ-muurolol had been determined in previous assays with RoseRS_3509, Rcas_0662 and SSCG_03688. The "cadinene" peak was made up of up to three products with slightly different retention times that could not be separated (Appendix Figure 7). These were all predicted to have a cadinene-like structure with δ-cadinene being predicted with the highest confidence. This assembly of peaks was seen in most samples and was determined as a single peak, named cadinene for this study. Any terpene products for which the NIST database could not accurately predict a unique structure were labelled as "other". Some of the predicted terpene structures are shown in Figure 5.13. All of

these terpenes are produced by a mechanism similar to τ-muurolol. Cubebol, *epi*-cubebol and β-copaene would require an extra cyclisation as well as another hydride shift for cubebol and *epi*-cubebol (Dickschat *et al.* 2014), respectively.

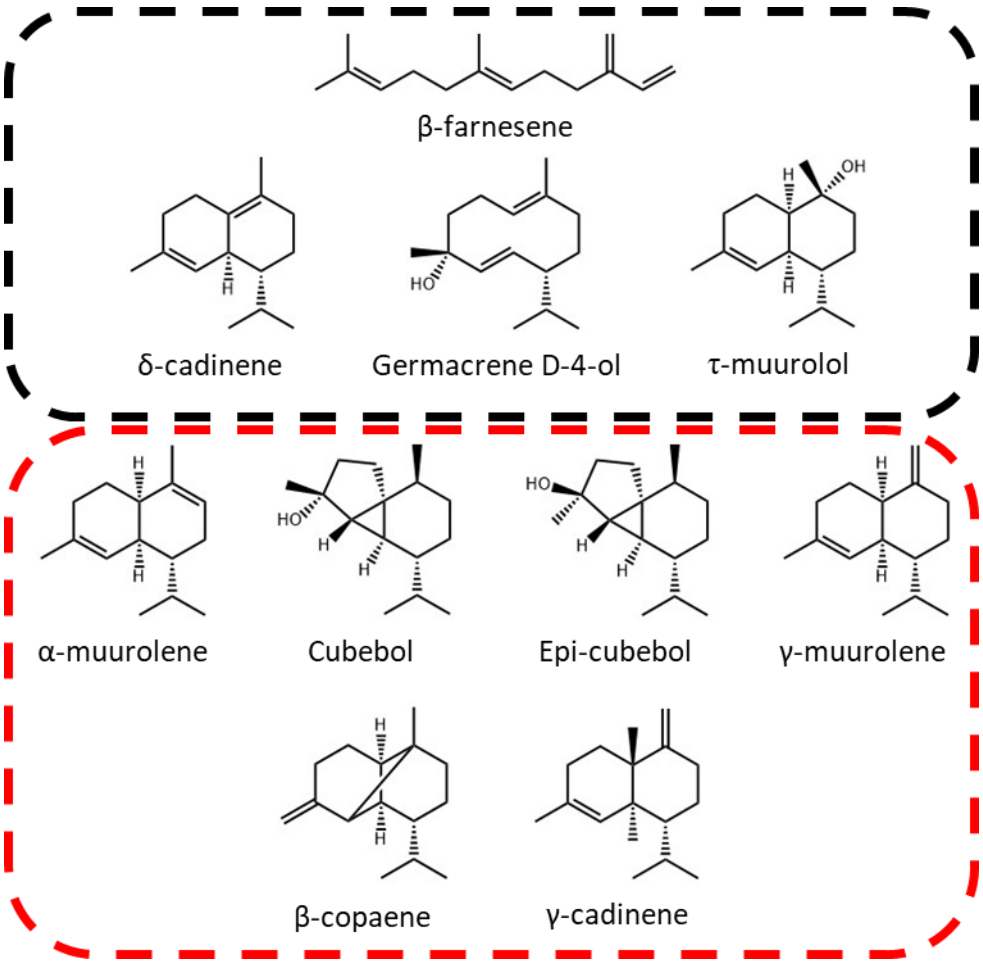


Figure 5.13: Terpene structures confirmed and predicted to be produced by the RoseRS_3509 active site mutants. In TPS product mixes, the terpene structures in the black box were those that have been previously confirmed or predicted by the NIST database with high confidence. Those in the red box are the terpenes which the NIST database predicts with lower confidence and therefore labelled as other terpene peaks in the XIC and FID data.

The FID and XIC results of the WT and each mutant at 37 °C are shown in Figure 5.14 and Table 5.2. From the FID, RoseRS_3509 was shown to produce mainly τ-muurolol with a small amount of cadinene and several other minor products predicted by the XIC. Y59A and L77A had similar profiles to RoseRS_3509, however several other terpene peaks were observed by FID suggesting that these mutations only minimally perturb the cyclisation mechanism as the majority of product remains as τ-

muurolol. I74A and C184A produced a higher quantity of τ-muurolol than RoseRS_3509, however with a higher amount of the same by-products than the WT. A higher turnover than the WT was also seen by Yoshikuni *et al.* (2006) when converting a δ-cadinene synthase to a germacrene D-4-ol synthase. L54A had τ-muurolol as the dominant product but with an increased amount of cadinene and other products being produced. Mutation of this residue clearly enables deviation from the main cyclisation mechanism, allowing other terpenes to form, especially cadinenes. The terpene profiles of N300A and W303A were similar to L54A however there was more variation in the production of τ-muurolol, both had levels of cadinene that were consistent with that produced by RoseRS_3509. Also, similarly to L54A, the product distributions of N300A and W303A contained a large quantity of other terpenes suggesting that the reduction in size of their respective side chains was interfering with the natural mechanism and allowing for alternate product formation. W303A also produced the monocyclic alcohol germacrene D-4-ol. This suggests that W303A provides access of a water molecule to the carbocation formed after the initial cyclisation. Unfortunately, as the structures of the other products could not be determined, it is difficult to predict how these mutations affect the mechanism. However, the change in product distribution indicates that these residues are important for τ-muurolol production. Similarly, Y310A also produced germacrene D-4-ol and a much smaller quantity of τ-muurolol while maintaining the amount of cadinene made by the WT, RoseRS_3509. It also did not produce any other terpenes suggesting that the aromatic residues in this region serve to exclude water which can otherwise rapidly trap the intermediate formed after initial cyclisation.

The positions that were mutated twice, F78 and W296, showed very different profiles as predicted. The mutation of F78 had a dramatic effect on the terpene product. By replacing phenylalanine with alanine, no terpene products were observed on the FID or XIC, while mutation to valine produced only a small quantity of τ-muurolol with no other terpenes detectable on the FID. The only mutants to produce β-farnesene at 37 °C were those with mutations to the W296 residue. While mutation to leucine still enabled τ-muurolol to be formed, β-farnesene appeared to be the dominant product alongside various other products including cadinene and germacrene D-4-ol. Placing alanine in this position compromised the mechanism so that only a small amount of τ-muurolol was produced with β-farnesene as the dominant product. The XIC also showed that W296A produced a small amount of cadinene but this was not detected by FID. Both F78 and W296 appear to be key for τ-muurolol cyclisation.

Bacterial TPSs can have different product profiles at different temperatures. For example, *epi*-isozizaene synthase produced 93% *epi*-isozizaene at 20 °C but 99% at 4 °C showing an increase in specificity when the temperature is dropped. However, when the TPS was assayed at 37 ºC, only 79% of the product was *epi*-isozizaene with β-farnesene comprising 5% (Aaron *et al.* 2010, Li *et al.* 2014). To determine whether a change in terpene profile occurs at a higher temperature, the mutants that showed a large deviation from the product profile of RoseRS_3509 at 37 °C were assayed at 50 °C for 1 hour (Table 5.3 and Figure 5.15). These were L54A, F78V/A, W296L/A and Y310A.

Table 5.2: Terpene compounds predicted by the NIST database using the product fragmentation pattern for each mutant assayed at 37 ºC and analysed by GC-MS. For the other peaks for which structures could not be accurately predicted, the number of peaks on the XIC is shown as the denominator of the fraction. The numerator of the other peaks is the number of peaks that could be measured on the FID.

Terpene	Mutants												
	WT	L54A	Y59A	I74A	L77A	F78A	F78V	C184A	W296A	W296L	N300A	W303A	Y310A
β-farnesene	N	N	N	N	N	N	N	N	Y	Y	N	N	N
Cadinene	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Germacrene D-4-ol	N	N	N	N	N	N	N	N	N	Y	N	Y	Y
τ-muurolol	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Other	0/3	4/7	3/3	1/2	2/7	0/0	0/1	2/2	0/1	2/4	4/6	3/5	1/4

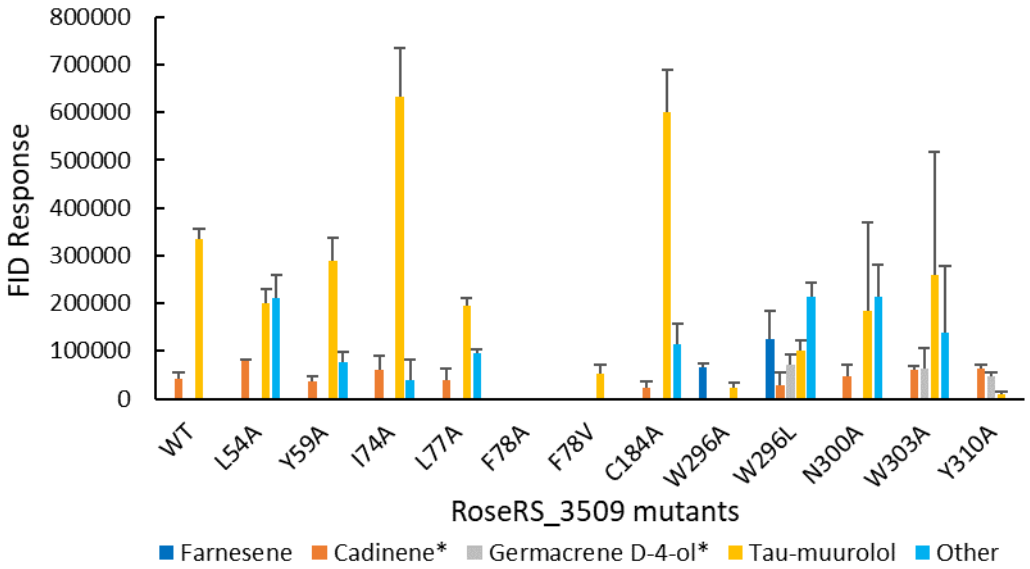


Figure 5.14: Products of active site mutants of RoseRS_3509 assayed at 37 ºC and quantified by GC-FID response. * The retention times of β-farnesene and τ-muurolol were known from previous assays but those of cadinene and germacrene D-4-ol could not be confirmed due to the lack of authentic standards.

At 50 ºC, unlike at 37 ºC, all of the mutants were shown to be active and able to turnover FPP (Table 5.3 and Figure 5.15), producing fewer other terpenes than were observed than at 37 ºC. The FID suggested that only τ-muurolol was produced at 50 °C by RoseRS_3509 although a small amount of cadinene and other terpenes were also produced in trace quantities only detectable by the XIC. L54A was the only mutant to produce less terpene overall at 50 °C than at 37 ºC, with the dominant product remaining as τ-muurolol with cadinene, germacrene D-4-ol and several other terpenes being formed as minor products. Both the XIC and FID show that Y310A only produced cadinene and germacrene D-4-ol at 50 °C with no τ-muurolol being detected. However, the main product of Y310A at 50 °C was germacrene D-4-ol rather than cadinene. From the XIC, F78V and W296L produced β-farnesene, cadinene, germacrene D-4-ol and τ-muurolol as well as some other terpenes. However, β-farnesene was not detectable by FID or XIC for F78V whereas it was one of the main products of W296L. F78V also produced a much smaller quantity of product than W296L. Of the alanine mutants at these two sites, only W296A produced β-farnesene as the dominant product while no products were detectable for F78A by XIC. As well as some τ-muurolol production by W296A was shown by the FID, W296A also produced detectable amounts of cadinene by XIC.

Table 5.3: Terpenes detected by the GC-MS XIC and predicted by the NIST database for RoseRS_3509 (WT) and its product distribution mutants assayed at 50 ºC. The other terpenes fraction was the number of undetermined structures detected by the FID over those detected by the XIC.

Terpene	Mutants						
	WT	L54A	F78A	F78V	W296A	W296L	Y310A
β-farnesene	N	N	Y	Y	Y	Y	N
Cadinene	Y	Y	Y	Y	Y	Y	Y
Germacrene D-4-ol	Y	Y	N	Y	Y	Y	Y
τ-muurolol	Y	Y	N	Y	Y	Y	N
Other	0/1	1/2	0/0	0/0	0/1	3/4	0/0

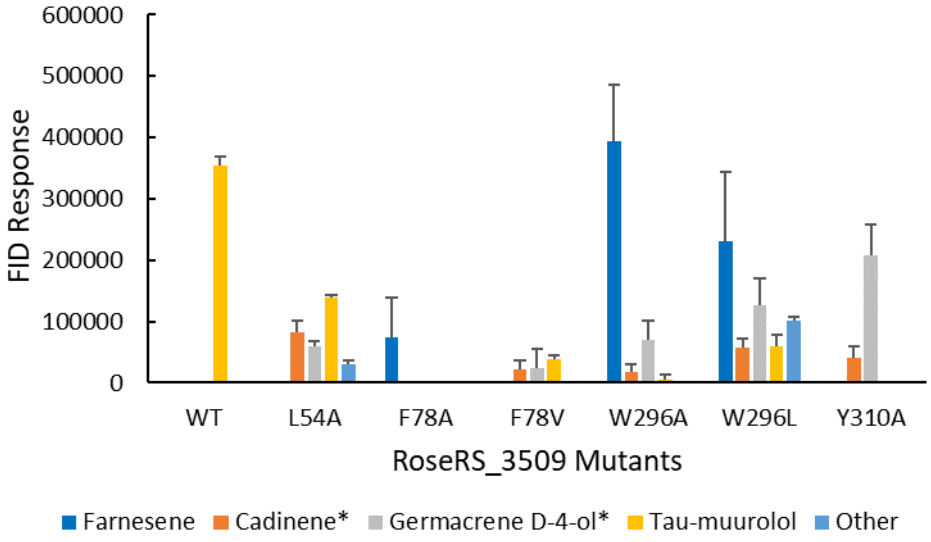


Figure 5.15: Abundance of the terpenes synthesised by RoseRS_3509 (WT) and the product distribution mutants using the GC-FID response at 50 ºC. An authentic standard was used for farnesene and the retention time of τ-muurolol was also known due to previous experiments. * Cadinene and germacrene D-4-ol were predicted by the NIST database.

Unlike the other mutant proteins tested at 50 ºC, L54A still produced τ-muurolol as its dominant product. Even though the side chain of leucine is small, the removal of the terminal isopropyl group still altered the product profile. While the amount of τ-muurolol produced by L54A decreased, this mutation led to the production of many unidentified other terpenes at 37 °C and germacrene D-4-ol at 50 ºC. L54 sits at the

base of the active site, where the introduction of the alanine mutation would provide only a small change in the cavity size. However, this mutation was enough to affect the binding and mechanism, underlining the importance of the shape of the active site contour. While most of the other mutants produced a higher quantity of terpenes at 50 ^{o}C, overall, L54A produced less than at 37 ^{o}C. This might suggest that successful substrate binding and turnover is less frequent in this mutant at higher temperatures compared to the WT and other mutants. Alternatively, as suggested by the CD melt curve, the secondary structure of L54A is unfolding at a higher rate at 50 °C than the other mutants which would result in less product produced over the extended assay time. This indicates that L54 may be more important to the overall stability of this enzyme compared to the other residues mutated here.

The lower turnover of Y310A at both 37 °C and 50 °C compared to the WT suggests that this mutation compromises binding, initiation or the progress of the reaction itself. As Y310 forms part of the RY motif, this is unsurprising as this motif aids coordination of the Mg^{2+} ions and the diphosphate of the substrate. In other enzymes, mutation of this residue to leucine left the resulting mutant inactive (Baer, Rabe, Fischer, *et al.* 2014). The predominant products at both temperatures were cadinene and germacrene D-4-ol. The increased quantity of germacrene D-4-ol from Y310A would suggest a premature termination of the reaction is occurring before the second cyclisation step. The increase in temperature seemed to prevent the second cyclisation from occurring as a smaller proportion of the cyclised product was reaching the cadinene or τ-muurolol stage at 50 °C than at 37 ^{o}C. The formation of germacrene D-4-ol indicates that the Y310A mutation enables access of a water molecule to the cation after the initial 1,10-cyclisation and 1,3-hydride shift. The hydroxyl is then added by nucleophilic attack of water to the monocyclic cation to form germacrene D-4-ol (Figure 5.16). The tyrosine residue might also be involved in stabilising the secondary carbocation formed after the initial cyclisation to enable the following 1,6-cyclisation to occur. This data therefore suggests that Y310A is a germacrene D-4-ol synthase. A crystal structure of the Y310A mutant of RoseRS_3509 with a substrate analogue could indicate the position of the water molecule responsible and the position of the tyrosine relative to the substrate during the reaction.

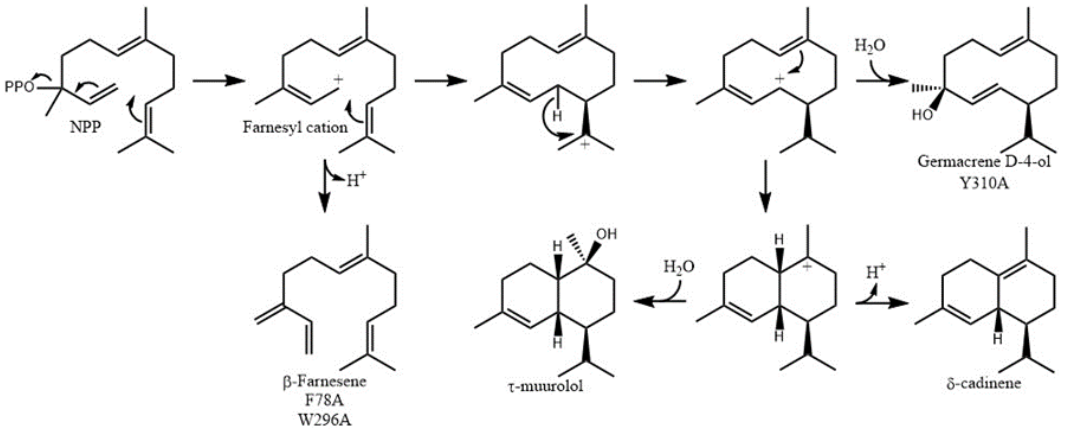


Figure 5.16: τ-muurolol cyclisation reaction of RoseRS_3509 as shown in Figure 5.7 but with the additions of the resulting active site mutants that produced β-farnesene and germacrene D-4-ol as dominant products. Image adapted from Dickschat *et al. (2014)*.

Combination of the 37 °C and 50 °C data suggests that the mutation of F78 and W296 had the largest change on the product profile of RoseRS_3509 as they both produced the linear β-farnesene. This demonstrates that the aromaticity and/or bulkiness of these residues is vital for the cyclisation mechanism as predicted from modelling. Both residues probably stabilise the carbocation intermediates through cation-π interactions with the aromatic side chains. The removal of these residues would therefore prevent the cyclisation mechanism from progressing towards τ-muurolol as the intermediate carbocations are not stabilised, resulting in alternative products being formed (Aaron *et al.* 2010). The dominant production of β-farnesene by the alanine mutants suggests that the mutation of these bulky residues enables the reaction to be terminated as soon as the diphosphate is removed. Both alanine mutants were able to make cyclised products indicating that FPP can still be bound in the native conformation and cyclised. However, for W296A, the mechanism is terminated after the first cyclisation to form germacrene D-4-ol. With the slightly larger aliphatic residues of valine and leucine, more FPP was converted to cyclised products but a significant fraction terminated immediately to form β-farnesene (Figure 5.16), suggesting that steric bulk alone, in the absence of aromaticity, can drive cyclisation. The majority of FPP binding in the alanine mutants is likely to be in an extended conformation similar to the situation described by Deligeorgopoulou and Allemann (2003). This may be the result of alternate FPP binding so that C10 or any other reactive carbon, usually C6 or C7, is not in close proximity to the carbocation formed at C1 by diphosphate removal. This would then allow for the carbocation to be

immediately terminated by deprotonation. However, without crystal structures of F78A and W296A with a substrate analogue, the absolute binding conformation of FPP in the active site is uncertain. As β-farnesene is the dominant product of F78A and W296A, enzymes carrying these mutations were characterised as β-farnesene synthases.

Only those mutants that contained a clear dominant product in the 50 °C assays were used for kinetic analysis at 50 °C; these were F78A, W296A and Y310A (Table 5.4). Kinetic analysis of RoseRS_3509 was also repeated as previous kinetic analysis by Styles *et al.* (2017) was carried out at 65 °C. At this temperature, the thermal melt curves clearly show that RoseRS_3509 and its protein mutants were unfolding so kinetics was performed at 50 °C as there was minimal unfolding at this temperature (Figure 5.12) Compared to the WT, W296A and Y310A had an increase in K_M while F78A showed a decrease. Mutation of the aromatic residues resulted in a decrease in k_{cat} by just over 2-fold for W296A with F78A and Y310A being a further 6-fold slower. Therefore, in terms of catalytic efficiency (k_{cat}/K_M), W296A only decreased ~3.5-fold compared to over 5-fold and 37-fold decreases for F78A and Y310A, respectively when compared to the WT.

Table 5.4: Kinetic parameters of RoseRS_3509 and its active site mutants with activities measured using the malachite green assay. Concentrations of FPP used were 0-46.2 μM for RoseRS_3509, F78A and Y310A and 0-92.4 μM for W296A. RoseRS_3509[1] is at 65 °C as shown in Styles *et al.* (2017). Michaelis-Menten kinetic curves of each enzyme are shown in Appendix Figure 18 except for RoseRS_3509 at 65 °C. Errors shown are standard errors.

Terpene Synthase	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($x10^{-3}$ $s^{-1}mM^{-1}$)
RoseRS_3509	9.8 ± 2.6	0.164 ± 0.016	16.8 ± 4.8
W296A	20.5 ± 4.0	0.078 ± 0.006	3.8 ± 0.8
F78A	5.2 ± 1.3	0.014 ± 0.001	2.7 ± 0.7
Y310A	23.6 ± 4.6	0.011 ± 0.001	0.4 ± 0.1
RoseRS_3509[1]	87.0 ± 31.0	0.950 ± 0.290	10.9 ± 5.9

Natural germacrene D-4-ol synthases are much more efficient than the Y310A mutant of RoseRS_3509 with the germacrene D-4-ol synthase from *S. citricolor* having a catalytic efficiency of 73.8 $mM^{-1}s^{-1}$ (Grundy *et al.* 2016). Y310A does however, have a much higher k_{cat} value than the germacrene D-4-ol synthases engineered by Yoshikuni *et al.* (2006) which were 0.0026 s^{-1} and 0.0019 s^{-1}. Those mutant

germacrene D-4-ol synthases were made using single mutations in the δ-cadinene synthase from the plant, *Gossypium arboreum*. As the catalytic efficiency of Y310A is lower than that of natural bacterial TPSs, it was likely to be too slow to be used in the *Parageobacillus* terpene platform. Nevertheless, Y310A is the first thermostable germacrene D-4-ol synthase to be characterised.

When compared to the other β-farnesene synthases produced by single mutations, the W296A mutant of RoseRS_3509 has the highest k_{cat}. The k_{cat} for F96A *epi*-isozizaene, W315P γ-humulene and Y92A aristolochene synthases were reported to be 0.0002 s^{-1}, 0.0019 s^{-1} and 0.0014 s^{-1}, respectively (Deligeorgopoulou and Allemann 2003, Yoshikuni, Martin, *et al.* 2006, Aaron *et al.* 2010). However, it is only second in terms of k_{cat}/K_M as the W315P γ-humulene synthase had a k_{cat}/K_M of 10.8 $mM^{-1}s^{-1}$, due to its incredibly low K_M value, 0.18 μM (Yoshikuni, Ferrin, *et al.* 2006). In nature, β-farnesene synthases are rare and have only been isolated from plants such as *A. annua* and *Mentha x piperita* (Crock *et al.* 1997, Salmon *et al.* 2015), however, the *M. piperita* β-farnesene synthase has not been characterised kinetically. The β-farnesene synthase from *A. annua* has a k_{cat}/K_M of 10.5 $mM^{-1}s^{-1}$ which is just over two-fold higher than W296A making this mutant almost as good as a WT β-farnesene synthase enzyme. W296A has a relatively similar k_{cat} to the TPS from *Thermostaphylospora chromogena* in chapter 3 as well as other bacterial TPSs reported in the literature which shows that it has activity equivalent to other bacterial TPSs (Nakano, Kim, *et al.* 2011a, Li *et al.* 2014, Styles *et al.* 2017). Thus, F78A and W296A are the first thermostable β-farnesene synthases to be identified with the latter having the faster turnover that is comparable to other bacterial TPSs.

Ideally, the next steps in this work would be to increase the turnover of F78A, W296A and Y310A. The highest throughput method for achieving this would be to use directed evolution with the colorimetric purpald assay to identify further mutations that increase the turnover towards β-farnesene or germacrene D-4-ol (Lauchli *et al.* 2013). To further understand the involvement of each of the residues in the mechanism of RoseRS_3509, each active site residue needs to be mutated to more amino acids than just alanine. Amino acids with different properties may help elucidate the role each residue has in terms of the mechanism. However, to truly understand the role of each residue, solved crystal structures of the WT and each mutant enzyme with a substrate analogue, such as DHFPP, are required. This would provide key information about the binding of FPP and the residues involved in each step of the reaction and allow for more informed decisions to be made for producing a specific terpene by mutation.

Adapting the strategy of previous work to produce β-farnesene by mutating aromatic residues at the base of the active site has enabled RoseRS_3509 to be converted into a β-farnesene synthase. This suggests that this strategy could be applied to any sesquiterpene synthase for conversion into a β-farnesene synthase if the correct residue, probably aromatic, is removed from the base of the active site. Future investigations might also include monoterpene and diterpene synthases for conversion to the simplest linear TPSs, namely geraniol and geranylgeraniol synthases using the same method of enabling extended binding conformations of the substrate. This would be of considerable interest when thermostable variants of these TPS types are characterised.

5.2.4 Farnesene production in *Parageobacillus*

The W296A mutant of RoseRS_3509 was chosen for introduction into the *Parageobacillus* terpene platform to generate a strain able to produce β-farnesene at high temperature. To enable the W296A mutant of RoseRS_3509 to be utilised in *Parageobacillus*, overlapping PCR was used to introduce the W296A mutation into the *Parageobacillus* codon optimised *roseRS_3509* sequence. The *Parageobacillus* optimised W296A (*geoopt_W296A*) mutant of *roseRS_3509* was inserted into the pG2K oriT SDM bgl sf-gfp GGready plasmid (Appendix Figure 9) behind a *prpls* promoter using Golden Gate cloning. The constitutive *prpls* promoters taken from Reeve *et al.* (2016) were *prplsWT*, *prpls1*, *prpls3*, *prpls5*, *prpls12* and *prpls18* which had a range of expression strengths in *E. coli* and *Parageobacillus*. Constructs containing the promoters *prpls1*, *prpls5* and *prpls18* were isolated and sequenced. The *prpls12 geoopt_W296A* construct contained the mutation, R301S, which was corrected using overlapping PCR. However, the correct constructs containing the strongest promoters in *E. coli*, *prplsWT* and *prpls3*, could not be made. Both had either single mutations, insertions, deletions or all three in either the *prpls* promoter or *geoopt_W296A* gene sequences. High constitutive expression of *geoopt_W296A* from the *prplsWT* and *prpls3* promoters may have caused cell stress, either because of the burden on the protein synthetic machinery or because the active enzyme could be depleting the cellular pool of FPP. Multiple attempts were made to rectify this problem by reducing possible toxicity. Firstly, transformed Golden Gate *prplsWT* constructs were recovered at RT and 30 °C instead of 37 °C. The lower temperature should reduce the amount of *geoopt_W296A* expressed (Zhang *et al.* 2003, de Groot and Ventura 2006, Rosano and Ceccarelli 2014) However, the sequenced constructs

still contained multiple mutations or indels. Next, the *prplsWT/3* and *geoopt_W296A* were inserted directly into pG2AC oriT SDM (Appendix Figure 9) using Gibson Assembly. These constructs were transformed into *E. coli* C43(DE3) but no colonies contained the successful construct even though the correct bands were seen from the Gibson Assembly for overlapping regions of the assembled plasmids (Figure 5.17). The colonies were tested for both the *prplsWT/3 geoopt_W296A* construct and the ampicillin resistance from pG2AC oriT SDM but no colonies contained either. As the toxicity could not be overcome, these constructs were discontinued.

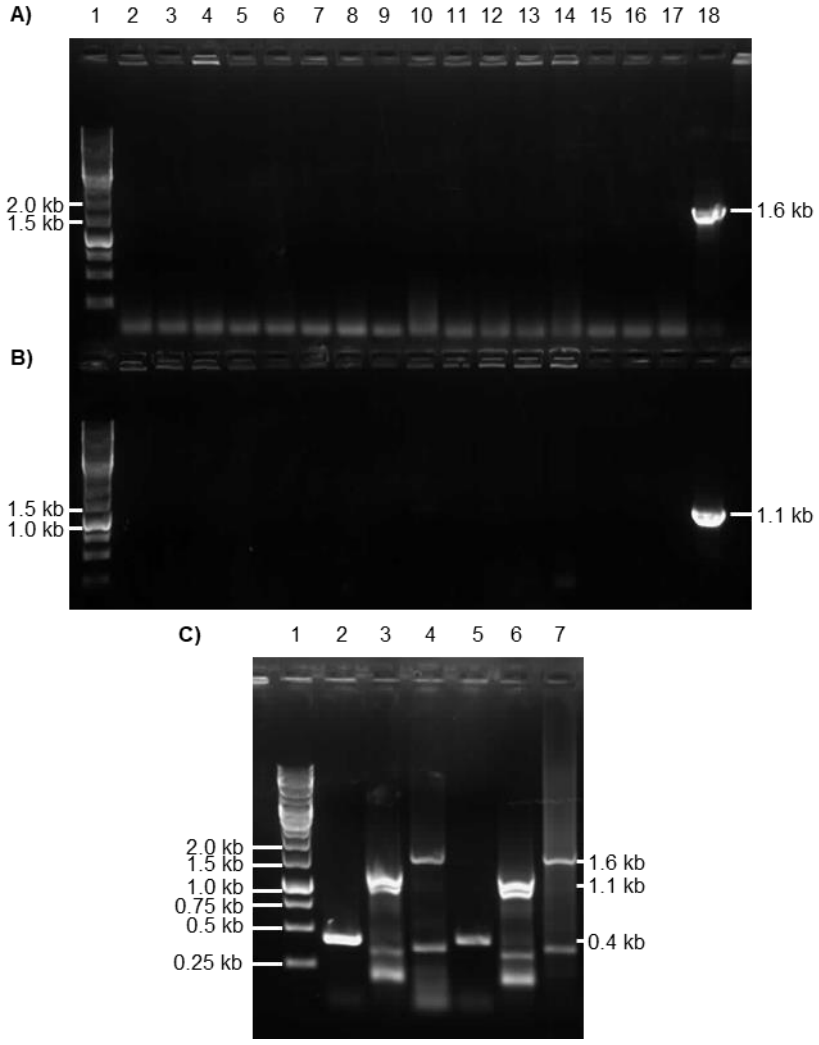


Figure 5.17: Diagnostic agarose gels for the cPCR of *E. coli* C43(DE3) colonies containing pG2AC oriT SDM *prplsWT/3* and the PCR amplification of the Gibson Assembly reactions. A) cPCR for the presence of the promoter-gene constructs of A) *prplsWT/3 geoopt_W296A* (1.6 kb) and B) ampicillin resistance gene (1.1 kb) of *E. coli* C43(DE3) that should contain pG2AC oriT SDM *prplsWT/3 geoopt_W296A*. Lanes 2-9 display products from 8 colonies potentially containing pG2AC oriT SDM *prplsWT geoopt_W296A* and lanes 10-17 are of 8 colonies that should contain pG2AC oriT SDM *prpls3 geoopt_W296A*. No bands were observed for each of the colonies tested. Lane 18 was the positive

control of the plasmid, pG2AC oriT SDM *prpls5 geoopt_W296A*, for the A) *prpls geoopt_W296A* construct and B) ampicillin resistance gene. C) PCR amplifications of the overlapping regions of the Gibson Assembly reactions used to make the pG2AC oriT SDM *prplsWT geoopt_W296A* (Lanes 2-4) and pG2AC oriT SDM *prpls3 geoopt_W296A* (Lanes 5-7) plasmids. Respective amplifications were for the overlap between the pG2AC oriT SDM backbone and the *prpls* promoter (0.4 kb) in lanes 2 and 5, the overlap between the *geoopt_W296A* gene and the pG2AC oriT SDM backbone (1.1 kb) in lanes 3 and 6 and the MCS of pG2AC oriT SDM that should contain the promoter-gene sequences of *prpls geoopt_W296A* (1.6 kb) in lanes 4 and 7.

The *prpls1/5/18 geoopt_W296A* fragments were restriction digested out of pG2K oriT SDM bgl sf-gfp GGready and ligated into pG2AC oriT SDM. The *prpls12 geoopt_W296A* overlapping PCR fragment was restriction digested and inserted into pG2AC oriT SDM. This vector was used because one of the strains that the constructs would be conjugated into, *Parageobacillus thermoglucosidasius* NCIMB 11955 cMev9, already contained a kanamycin resistance marker, therefore, a vector containing the only other antibiotic resistance marker that could be used in *Parageobacillus*, chloramphenicol, was required. The restriction enzyme pairs were NheI/SmaI for *prpls1 geoopt_W296A* while NheI/HindIII was used for *prpls5/12 geoopt_W296A*. Unfortunately, the pG2AC oriT SDM *prpls18 geoopt_W296A* plasmid could not be constructed using the NheI/HindIII pair. Although positive colonies were identified by cPCR, each construct contained deletions of between 5-18 bp. Attempts to make this construct by Gibson assembly were also unsuccessful, as they contained deletions of varying lengths that could not be corrected by Quikchange site directed mutagenesis. As *prpls18* only expresses constitutively at very low levels in *Parageobacillus* compared to the successfully made constructs, it was not pursued further.

The pG2AC oriT SDM *prpls1/5/12 geoopt_W296A* constructs were transformed into *E. coli* S17.1 and conjugated into *P. thermoglucosidasius* NCIMB 11955 and *P. thermoglucosidasius* NCIMB 11955 cMev9. For *P. thermoglucosidasius* NCIMB 11955, production of β-farnesene from GeoOpt_W296A needed to use the native FPP pool while it was hoped that *geoopt_W296A* would be expressed at a higher level than *roseRS_3509* in the chromosomal Mev9 pathway of *P. thermoglucosidasius* NCIMB 11955 cMev9. This would mean that both β-farnesene and τ-muurolol would be produced by the cMev9 strain as GeoOpt_W296A competes with RoseRS_3509 for FPP. Multiple plasmid containing colonies were confirmed by cPCR for each construct (Figure 5.18), although pG2AC oriT SDM *prpls5 geoopt_W296A* could not be conjugated into the cMev9 strain. Cultures of each strain

in 50 ml 2YT culture containing the appropriate antibiotics were incubated at 50 °C with a 5 ml dodecane top-layer, and the cMev9 strains were induced with 1% maltose at an OD_{600} of 0.1. Samples (15 µl) of the dodecane layer were taken after 48 hours and diluted 1 in 10 in hexane before running on the GC-MS. However, no β-farnesene peak was observed on the spectra, whereas τ-muurolol was observed on the cMev9 spectra. This was also attempted in 10 ml culture volumes also but with the same result. This might be due to a lower level of GeoOpt_W296A production in *P. thermoglucosidasius* NCIMB 11955 cMev9 than expected which meant that only τ-muurolol was produced and the native FPP pool in *P. thermoglucosidasius* NCIMB 11955 being too small to allow β-farnesene production to reach a level that could be detected by GC-MS. To fully test this system in the future, RT-qPCR could be used to determine if *geoopt_W296A* is expressed in both of these strains and then optimise conditions to enable β-farnesene to be produced.

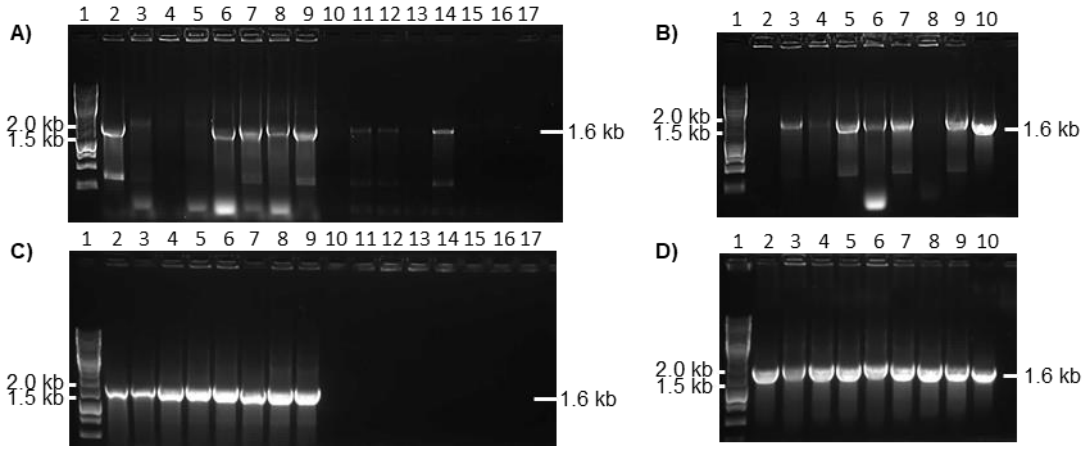


Figure 5.18: cPCR amplifications of the promoter-gene constructs of *prpls geoopt_W296A* sequences together from *P. thermoglucosidasius* NCIMB 11955 (A and B) and *P. thermoglucosidasius* NCIMB 11955 cMev9 (C and D) colonies that should contain pG2AC oriT SDM *prpls1/5/12 geoopt_W296A* constructs. The 1.6 kb amplicon was the *prpls* promoter and *geoopt_W296A* gene together. Lane 1 in A-D was the ladder, GeneRuler 1kb DNA Ladder. A) cPCR of colonies 1-8 for the *prpls1 geoopt_W296A* construct (Lanes 2-9) and the *prpls5 geoopt_W296A* construct (Lanes 10-17) from *P. thermoglucosidasius* NCIMB 11955. B) cPCR of colonies 1-8 for the *prpls12 geoopt_W296A* construct (Lanes 2-9). C and D are 8 colonies screened in the same lanes as A and B but from *P. thermoglucosidasius* NCIMB cMev9 colonies that should contain the pG2AC oriT SDM *prpls1/5/12 geoopt_W296A* constructs also.

An alternative strategy would be to construct the full Mev9 pathway but with *geoopt_W296A* in place of *roseRS_3509*. Two methods developed for this purpose.

The first method was to build the pathway using Golden Gate cloning and restriction digestion-ligation to insert the promoter and *geoopt_W296A* into the vectors made by Styles *et al.* (To be published) to create the plasmid, P2 Mev9. The second method was to knock-out *roseRS_3509* from the *P. thermoglucosidasius* cMev9 NCIMB 11955 strain and knock-in *geoopt_W296A* in its place. Both methods also use the maltose-inducible promoter, *pglv*, instead of the constitutive *prpls* promoters so that *geoopt_W296A* would only be expressed at the same time as the rest of the pathway by using maltose. For the first strategy, the *pglv* promoter and *geoopt_W296A* were inserted into P2 UM3 by Golden Gate cloning. The P2 Mev9 plasmid containing *geoopt_W296A* was constructed by restriction digestion-ligation of P2 LM5, containing a mutation in diphosphomevalonate decarboxylase (DMD) to remove an EcoRI site, and P2 UM3 *geoopt_W296A*. Both vectors were restriction digested with SacI but the second digestion of P2 LM5 was with XbaI while P2 UM3 *geoopt_W296A* was with BcuI. The XbaI-LM5-SacI was inserted into P2 UM3 *geoopt_W296A* as XbaI and BcuI have the same 4 bp overhangs. The overlapping region of the plasmid was sequenced to show that it contained both UM3 and LM5. However, the next steps require complete sequencing of the whole Mev9 pathway before introducing the attP site for insertion into the *P. thermoglucosidasius* NCIMB 11955 genome using the serine integrase system (Styles *et al.*, to be published). After these checks, testing for β-farnesene production can begin.

The mutation of RoseRS_3509 to a β-farnesene synthase, W296A, requires only a 2 bp change in the gene from the tryptophan codon, TGG, to the alanine codon, GCG. However, this change would be incredibly difficult to do in the *P. thermoglucosidasius* NCIMB 11955 cMev9 strain by homologous recombination as the 2 bp mutation may not be incorporated into the genome if the crossovers occur in distal positions. To identify a mutant, a high number of colonies would need to be screened by sequencing of PCR products which would be inefficient and time-consuming.

Knocking out *roseRS_3509* from *P. thermoglucosidasius* NCIMB 11955 cMev9 using simple homologous recombination of *geoopt_W296A* and *roseRS_3509* would be challenging as only 2 bps need to be changed which would require a large number of colonies to be screened in order to find a successful mutant. One optimal alternative was to knock-out *roseRS_3509* and then replace it with *geoopt_W296A* with the knock-in vector. The knock-out vector would replace the kanamycin resistance gene and *roseRS_3509* with the chloramphenicol resistance gene using homologous recombination between the repB sequence and a partial *hmgr* sequence. The knock-in vector would then reinsert the kanamycin resistance and *geoopt_W296A* in the

same position with the chloramphenicol removed. Both vectors are shown in Figure 5.19. While both of these vectors were constructed, the next steps for performing the knock-out and knock-in need to be performed in the future.

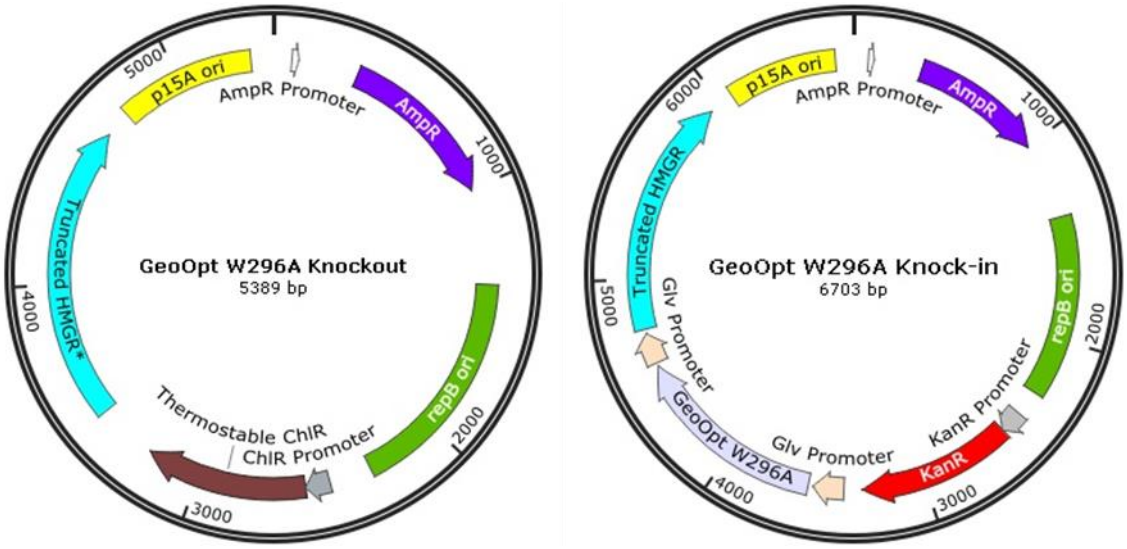


Figure 5.19: Knock-out and knock-in vectors for removing *roseRS_3509* from *P. thermoglucosidasius* NCIMB 11955 cMev9 and replacing with *geoopt_W296A*. The homologous recombination regions are in the repB origin and truncated *hmgr* sequence. *The truncated *hmgr* gene in the knock-out vector lacked the first 19 bp.

5.3 Conclusions

TPSs can be promiscuous enzymes with the ability to produce multiple terpenes from a single cyclisation mechanism. This promiscuity allows the mechanism of these enzymes to be altered by manipulating the active site contour resulting in a different dominant product from that produced by the WT (Yoshikuni, Martin, *et al.* 2006, O'Maille *et al.* 2008, Aaron *et al.* 2010, Li *et al.* 2014). Several groups have targeted production of the linear terpene, β-farnesene, by expanding the TPS active site to allow for extended binding of the substrate FPP. Although β-farnesene was a dominant product in these studies, all enzymes lost much of their activity (Deligeorgopoulou and Allemann 2003, Yoshikuni, Ferrin, *et al.* 2006, Aaron *et al.* 2010). This study planned to produce the first thermostable β-farnesene synthase by applying a similar approach to a thermostable TPS as it had a superior k_{cat} compared to the TPSs altered previously. Therefore, this could produce a β-farnesene synthase that maintains a higher turnover.

The structure of the thermostable τ-muurolol synthase, RoseRS_3509, from *Roseiflexus* sp. RS-1 was modelled to establish the residues that formed the active site cavity. Of these, the hydrophobic residues were mutated to alanine individually to determine which residues were vital for catalysis. To increase the chances of producing β-farnesene, aromatic residues towards the base of the active site were targeted, especially F78 and W296 which were mutated to the shorter aliphatic side chains of valine and leucine, respectively. Mutants were tested for altered stability by CD where most maintained similar stabilities to RoseRS_3509 except for L54A and L77A. All mutants were tested at 37 °C using the GC-MS enzyme assay to determine the deviation in the product profiles from the WT. Only Y59A maintained a similar profile to the WT. All of the other mutants showed a different product profile suggesting that each residue was involved in the mechanism. In most cases, there was a decrease in the proportion of τ-muurolol in the products with other terpenes such as cadinene, germacrene D-4-ol and β-farnesene being produced. Only W296A/L produced β-farnesene at 37 °C where F78V produced a small amount of τ-muurolol and F78A seemed to be inactive. F78A/V, W296A/L, L54A and Y310A were selected to be tested at 50 °C. L54A showed a decrease in terpene product, probably resulting from decreased structural thermostability caused by the mutation. The dominant product of Y310A at 50 °C was germacrene D-4-ol suggesting that this residue is important for stabilising the carbocation after the first cyclisation and this rearrangement is terminated by hydroxylation. F78V and W296L produced a wide spread of terpene structures, but most significantly the production of β-farnesene alongside τ-muurolol, cadinene and germacrene D-4-ol. This indicated that these mutants were still able to cyclise FPP to τ-muurolol but, without the aromatic residues, intermediates were being quenched more frequently, giving rise to less complex terpene structures, thus lowering the specificity of this enzyme. F78A and W296A both produced β-farnesene as the dominant product suggesting that both of these residues are not only important for carbocation stabilisation but also the folding of FPP upon binding in the active site. The truncation of these residues results in alternate binding conformations that allow immediate termination of the farnesyl cation by deprotonation. The kinetics of F78A, W296A and Y310A were measured to produce the first characterised thermostable β-farnesene synthases and first engineered thermostable germacrene D-4-ol synthase. The sequence encoding for the faster of the two β-farnesene synthases, W296A, was inserted into *P. thermoglucosidasius* NCIMB 11955 with and without the chromosomal mevalonate pathway. However, no β-farnesene production was detected. Further strategies for detecting β-farnesene production in *Parageobacillus* species have been developed

including building a new mevalonate pathway to include W296A but also a knock-out/knock-in method to mutate *roseRS_3509* to *geoopt_W296A* in the *P. thermoglucosidasius* NCIMB 11955 strain.

Prior to this work, only three thermostable TPSs had been characterised, one of which was identified in chapter 4. These are the two τ-muurolol synthases from *Roseiflexus* species and the germacrene D-4-ol synthase from *T. chromogena*, respectively (Styles *et al.* 2017). This work was able to determine the hydrophobic residues involved in catalysis of τ-muurolol as well as identify aromatic residues, that upon mutation to alanine, created the first thermostable β-farnesene synthases as well as the first engineered thermostable germacrene D-4-ol synthase through model-guided engineering. This demonstrates that it should be possible to mutate any sesquiterpene synthase into the simplest sesquiterpene synthase, a β-farnesene synthase, when the appropriate residue, most likely aromatic, is removed from the base of the active site.

The development of a thermostable β-farnesene synthase provides a starting point for producing more terpenes at high temperatures. With more work to understand the residues involved in catalysis, more terpenes could be made from the RoseRS_3509 scaffold. This could include some of the cyclic hydrocarbons including the cadinenes or γ-muurolene. Henceforward, F78A, W296A and Y310A could be put through cycles of directed evolution using the purpald assay to engineer more catalytically efficient and robust TPSs (Lauchli *et al.* 2013). This should enable their use in the *Parageobacillus* terpene platform for possible future industrial application.

6 Discussion and Further Work

The development of microbial platforms in *E. coli*, *S. cerevisiae* and *R. sphaeroides* has enabled large scale production of valuable terpenes from sugar where only low yielding plant extraction methods or chemical synthesis were used previously. While extensive research has gone into the development of these systems, they use a food-stuff as a feedstock, where a waste feedstock may be a more sustainable solution if high terpene yields can be obtained. Styles *et al.* (To be published) developed the first terpene platform to function at high temperatures in *Parageobacillus thermoglucosidasius* NCIMB 11955. The ultimate goal of this platform is for valuable terpenes to be produced using lignocellulosic biomass as a waste feedstock. As well as the advantages of performing fermentations at high temperatures, this platform would be more sustainable and would not compete with the food market for the feedstock. Prior to this study, limited work had been carried out into the thermostability of TPSs and their properties. In fact, only two naturally thermostable TPSs had been characterised, both τ-muurolol synthases, from *Roseiflexus* species (Yamada *et al.* 2015, Styles *et al.* 2017). This initially restricted the *Parageobacillus* platform to only τ-muurolol production. The aim of this work was to therefore identify or generate novel thermostable TPSs to expand the potential applications of the *Parageobacillus* terpene platform. To do this, three strategies were identified. These strategies were to identify novel thermostable TPSs using bioinformatic search techniques and characterise these, to increase the thermostability of a mesophilic TPS using comparative modelling and computational prediction software followed by production of the TPS and, lastly, to alter the product distribution of a thermostable TPS to another terpene product.

Identifying novel naturally thermostable TPSs requires more complex bioinformatic search methods than BLAST analysis due to the very low sequence identity of bacterial TPSs, with the exception of several short, conserved motifs. Yamada *et al*. (2012, 2015) and Komatsu *et al*. (2008) showed that iterative hidden Markov models (HMMs) can be used to identify novel class I and II TPSs with low sequence identities from bacteria. These searches start with an initial search alignment, using this to identify new sequences based on a scoring system built around the alignment. The diversity of the sequences retrieved by the HMM is dependent on the sequences used in the multiple sequence alignment input suggesting that although more similar sequences will be identified, some sequences can still be missed. To sample the

increasing number of bacterial genomes for TPSs from thermophiles, two separate HMMs were performed; one for class I TPSs and the other for class II TPSs. The sequences selected as the input for each HMM were not only chosen for their low sequence identities but also for their variations in the conserved motifs to increase the number of TPS sequences that were identified. From both HMMs, novel TPS sequences were identified from thermophilic organisms but none from hyperthermophiles. For the sequences identified from the class I HMM, several of the putative thermostable TPSs had been fully characterised previously, τ-muurolol synthases (RoseRS_3509 and Rcas_0662) from *Roseiflexus* species, and for others just their terpene products, e.g. cembrene C synthase (Rxyl_0493) from *R. xylanophilus* while the sequence from *T. curvata* (Tcur_3107) had been identified but no activity was seen in *in vivo* assays (Yamada *et al.* 2015, Styles *et al.* 2017). Of these TPSs, only RoseRS_3509 was used in the initial sequence alignment. Novel bacterial TPS sequences were identified from *A. eurytherma*, *F. thermalis*, *T. bispora* DSM 43183 and *T. chromogena*. Most of the products for these suggested TPSs could not be predicted from their sequences except EDD35_0439 from *A. eurytherma* which was predicted to be a geosmin synthase in this study. Another sequence from the thermophile *K. papyrolyticum* JKG1, was identified as a class I TPS by a BLAST search with the thermostable RoseRS_3509 in this study. Of the sequences identified by the class II HMM, most represented squalene cyclases rather than diterpene synthases. However, some of the sequences that were predicted to be class II diterpene synthases from thermophiles did not have a class I partner in the genome for the removal of the diphosphate. It is currently unclear whether these were gained by horizontal gene transfer without a class I TPS partner, the cyclised product is utilised by another enzyme instead of a class I TPS or whether the class I partner just hasn't been identified, therefore more investigation is needed. Of the novel class I TPSs, four were chosen for further characterisation. These were the sequences from *F. thermalis* (*FJSC11*), *K. papyrolyticum* JKG1 (*JKG1*), *T. bispora* DSM 43183 (*Tbis_3257*) and *T. chromogena* (*Tchrom*). The previously identified TPSs from *T. curvata* (*Tcur_3107*) and *R. xylanophilus* (*Rxyl_0493*) were also selected for thermostability experiments while the former was also tested for *in vitro* activity as previous studies have only tested it *in vivo* (Yamada *et al.* 2015). Most of the proteins tested by CD were mildly thermostable but the putative TPSs from *T. bispora* and *T. chromogena* were characterised as thermostable with T_m values of 78.5 ºC and 57.1 ºC, respectively. While all of the purified enzymes showed activity with GPP and FPP, except Tcur_3107, most were predicted to be sesquiterpene synthases due to the higher turnover. From the *in vitro* GC-MS assays, JKG1, FJSC11 and Tchrom were

predicted to be hedycaryol, (+)-sativene and germacrene D-4-ol synthases, respectively. If these structures have been correctly predicted, then this study has identified the first (+)-sativene synthase and the first naturally thermostable germacrene D-4-ol synthase but NMR analysis is required to confirm these structures. To do this analysis, an engineered system is required to produce higher quantities than is required for an *in vitro* enzyme assay. For example, the engineered *Streptomyces avermitilis* strain SUKA22 which has been optimised for secondary metabolite biosynthesis by having all other secondary metabolite pathways or genes deleted to remove competing carbon sinks could be utilised (Komatsu *et al.* 2013, Yamada *et al.* 2015). While Tbis_3257 was able to turn over GPP and FPP, it was predicted to be a diterpene synthase. The bacterial class II TPS HMM identified *Tbis_3256*, which sits next to *Tbis_3257* in the genome. As Tbis_3257 was shown to be highly thermostable, further investigation to determine whether this pair of TPSs work together to make a diterpene product at high temperatures would be valuable. If so, this class I/II TPS pair could be suitable for use in the *Parageobacillus* platform. Further work is also required to characterise the other putative TPSs identified in this study, especially those from *A. eurytherma* and the class I/II pair from *T. curvata*, Tcur_3806/Tcur_3808, as functional TPSs have not been identified from these two organisms before. This study has shown that running HMMs with different multiple sequence alignment inputs can identify novel thermostable TPSs and this might suggest that more thermostable TPSs may be identified from running more HMMs with different initial inputs. As only three naturally thermostable TPSs have been characterised currently from this work and Styles *et al*. (2017), it is possible that class I TPSs are not as widely utilised at high temperatures as they are in mesophilic bacteria. This may be due to the increased volatility of small terpenes at high temperatures.

The second strategy used to generate more thermostable TPSs was to increase the thermostability of a mesostable TPS. Initially, a homology model comparison was performed between a mesostable τ-muurolol synthase, SSCG_03688, and the thermostable τ-muurolol synthases, RoseRS_3509 and Rcas_0662, to determine if there were any differences in the secondary structure. Although these TPSs were highly structurally similar, SSCG_03688 had C-terminal extension of >90 aa. The sequential truncation of this tail not only increased the thermostability of the τ-muurolol synthase by up to ~1.2 ^{o}C but also identified the limit of C-terminal truncation before stability of the protein was affected. The most thermostable τ-muurolol synthase truncation protein, called SSCG_R6, was truncated by 57 aa. Only few of

studies have previously attempted to increase the thermostability of a TPS. Rounds of directed evolution have been shown to successfully increase TPS thermostability by >10 ^{o}C without affecting catalytic activity but this approach relies on screening a large number of mutants (Lauchli *et al.* 2013). As well as visually comparing differences in secondary structure, there are a number of computational programmes that have been developed to predict thermostabilising mutations in individual proteins. Previously, a study into 5-*epi*-aristolochene synthase used a computational method to predict thermostabilising mutations and consequently introduced twelve mutations in a single protein resulting in a thermostable TPS that was almost completely inactive (Diaz *et al.* 2011). During this thesis work, two computational methods, HoTMuSiC and Rosetta, were used separately with a homology model to determine mutations that would stabilise the most thermostable C-terminally truncated τ-muurolol synthase developed previously, SSCG_R6. Of the seven mutations selected from the HoTMuSiC prediction, four resulted in insoluble protein on overexpression. Of the soluble mutant proteins generated and isolated from this work, only a single mutation, G24I, increased the thermostability in terms of the structure and activity compared to SSCG_R6. As well as the SSCG_R6 homology model, the crystal structure of selinadiene synthase (SdS, PDB accession code 4OKM) was also inserted into the Rosetta server to predict thermostabilising mutations. Five predicted mutations were picked in the same position of the SSCG_R6 homology model and SdS structure. Of the SSCG_R6 protein mutations, one was insoluble while the rest showed little to no change in thermostability. On the other hand, all of the SdS mutations were soluble but had more variation in terms of thermostability; G45K and T273K increased the T_m by ~1.0 ^{o}C, while the other three mutations resulted in a decrease of up to ~3.6 ^{o}C in T_m. For more accurate predictions of thermostabilising mutations, most prediction methods recommend a crystal structure as an input, but this was not available for SSCG_R6 (Yu and Huang 2013). As a homology model is predicted from another protein sequence with high similarity or identity, it will likely contain some errors and therefore may lead to falsely predicted thermostabilising mutations. While thermostabilising mutations for the τ-muurolol and selinadiene synthases were identified in this work, none increased the stability to a high enough extent for either TPS to be used in the *Parageobacillus* platform. This work suggests that the high throughput directed evolution method remains the more successful method for increasing TPS thermostability, even considering the limitations discussed above, rather than using prediction programs that are much lower throughput and rely on the availability of crystal structures for more accurate TPS thermostabilising mutation prediction.

The final strategy used to increase the potential applications of the thermostable *Parageobacillus* terpene production platform was to change the product distribution of a known thermostable TPS. In this study, the active site of the thermostable τ-muurolol synthase, RoseRS_3509, was manipulated to shift the specificity towards the production of β-farnesene. Studies by Aaron *et al*. (2010), Deligeorgopoulou *et al.* (2003) and Yoshikuni *et al.* (2006) have shown that altering the active site cavity of a TPS in a way that changes the substrate binding and the cyclisation mechanism can change the resulting product. All three of these studies were able to identify a single mutation that converted the *epi*-isozizaene, aristolochene and γ-humulene synthases into β-farnesene synthases. It was predicted that removing an aromatic residue at the base of the active site would enable FPP to bind in an extended conformation that would allow for immediate termination by deprotonation after reaction initiation to create β-farnesene. The aim of chapter 5 was to produce the first thermostable β-farnesene synthase by targeting aromatic residues in the active site, especially those at the base of the active site, of the thermostable τ-muurolol synthase, RoseRS_3509. As no crystal structure could be produced of RoseRS_3509, a homology model had to be used to identify target residues. Other residues predicted to be in the lower hydrophobic part of the active site were also mutated to determine their involvement in τ-muurolol cyclisation. Most of the amino acids that were mutated affected the product distribution at 37 ^{o}C. The mutants W296A and Y310A were shown to produce β-farnesene and germacrene D-4-ol, respectively, suggesting that they were allowing the carbocation to be quenched at different points in the mechanism. The GC-MS assays of the aromatic mutants at 50 ^{o}C showed that F78A and W296A were β-farnesene synthases and Y310A was a germacrene D-4-ol synthase while F78V and W296L each produced a variety of terpene products from different intermediate carbocations. This suggested that mutations to alanine at positions F78 and W296 allowed FPP to bind in an alternative conformation that prevents cyclisation while using larger aliphatics or aromatics in these positions allows for more reactions to form cyclic products such as germacrene D-4-ol, cadinene or τ-muurolol. The Y310A mutation was predicted to enable the carbocation to be quenched after the initial cyclisation to form the monocyclic, germacrene D-4-ol. In combination with the other studies, this work suggests that almost any class I TPS has the ability to produce the simplest sesquiterpene, β-farnesene, through a single mutation, at the base of the active site, usually of an aromatic amino acid. However, mutations of the active site tend to have a slight negative impact on the catalytic activity which was seen in most of the previous studies and in the mutant proteins made here. Nevertheless, the W296A mutation had similar kinetic constants to other bacterial TPSs and was

inserted into the *Parageobacillus* platform to produce β-farnesene. So far, no β-farnesene has been produced *in vivo* but more work is required to enable β-farnesene to be produced at high temperatures for the first time. This might include improving the turnover of W296A using the directed evolution method developed by Lauchli *et al.* (2013). The turnover might prove to be a limitation of the thermostable β-farnesene synthase generated in this work. This method could also be used to improve turnover of F78A and Y310A for β-farnesene and germacrene D-4-ol, respectively, so that they may also be used in the *Parageobacillus* platform. This work has begun the identification of active site amino acids involved in the cyclisation mechanism of RoseRS_3509 and how they affect the mechanism. To determine how each residue is involved in the cyclisation, details of how the substrate binds in the active site, the interaction between each amino acid and the carbocation intermediates and how active site mutations shift the product distribution need to be determined. This would require crystal structures with substrate analogues to show how FPP initially binds in the active site. In addition, a more extensive study of how each amino acid is involved in the mechanism could be conducted by introducing further mutations to active site residues and determining how these mutations affect the mechanism. To introduce these mutations, iterative saturation mutagenesis of the active site amino acids could be used. This would also enable the position of the water molecule that quenches the carbocation in τ-muurolol cyclisation to be identified. Identifying the position of this water molecule would allow for mutations to be made that prevent the addition of the hydroxyl and allow for bicyclic terpene hydrocarbons to be produced instead.

Of the three strategies tested in this work, the bioinformatic approach and altering the product distribution of RoseRS_3509 proved the most successful. These methods increased the number of terpenes that could be produced at high temperatures *in vitro* from one to three. The HMMs run here as part of the bioinformatic approach and previous work by Yamada *et al.* (2015) suggest that there are fewer thermostable TPSs than mesostable variants in bacteria. Therefore, engineering TPSs by either introducing thermostabilising mutations into mesostable TPSs or altering the product distribution of a thermostable TPS may be more efficient approaches towards producing more thermostable bacterial TPSs (Dickschat 2016). It is also worth noting that the majority of industrially valuable terpenes are currently isolated from plants and therefore future research could target plant class I TPSs for use in the *Parageobacillus* terpene platform. All of the commercial microbial terpene platforms currently use plant TPSs for the production of valuable terpenes. As no plant TPSs have been characterised as naturally thermostable, the thermostability of both the α-

domain and the class II domains that are present would need to be increased. While prediction software to do this could be used, the work in this thesis and the current literature suggests that using directed evolution may be more successful for increasing the thermostability of a TPS as many plant TPSs do not have crystal structures, which may limit the accuracy of predictions of thermostabilising mutations (Diaz *et al.* 2011, Lauchli *et al.* 2013). A major limitation of this method is the number of mutants required to be screened for each individual TPS. Engineering more thermostable TPSs such as those from plants would enable other industrially valuable terpenes, such as isoprene, amorphadiene and valencene, to be produced at high temperatures, which would present a great advantage to many industries (Leavell *et al.* 2016). This work provides an initial starting point for determining which strategies are better for identifying or making thermostable TPSs to permit the *Parageobacillus* terpene platform to be used for these purposes.

7 References

Aaron, J.A., Lin, X., Cane, D.E., and Christianson, D.W., 2010. Structure of Epi-Isozizaene Synthase from Streptomyces coelicolor A3(2), a Platform for New Terpenoid Cyclization Templates. *Biochemistry*, 49(8), 1787–1797.

Abdulla Bava, K., Gromiha, M.M., Uedaira, H., Kitajima, K., and Sarai, A., 2004. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Research*, 32, D120–D121.

Acevedo-Rocha, C.G., Gronenberg, L.S., Mack, M., Commichau, F.M., and Genee, H.J., 2019. Microbial cell factories for the sustainable manufacturing of B vitamins. *Current Opinion in Biotechnology*, 56, 18–29.

Adimpong, D.B., Sørensen, K.I., Nielsen, D.S., Thorsen, L., Rasmussen, T.B., Derkx, P.M.F., and Jespersen, L., 2013. Draft Whole-Genome Sequence of Bacillus sonorensis Strain L12, a Source of Nonribosomal Lipopeptides. *Genome Announcements*, 1(2), 1–2.

Agger, S., Lopez-Gallego, F., and Schmidt-Dannert, C., 2009. Diversity of sesquiterpene synthases in the basidiomycete Coprinus cinereus. *Molecular Microbiology*, 72(5), 1181–1195.

Agger, S.A., Lopez-Gallego, F., Hoye, T.R., and Schmidt-Dannert, C., 2008. Identification of Sesquiterpene Synthases from Nostoc punctiforme PCC 73102 and Nostoc sp. Strain PCC 7120. *Journal of Bacteriology*, 190(18), 6084–6096.

Aharoni, A., Giri, A.P., Deuerlein, S., Griepink, F., de Kogel, W.-J., Verstappen, F.W.A., Verhoeven, H.A., Jongsma, M.A., Schwab, W., and Bouwmeester, H.J., 2003. Terpenoid Metabolism in Wild-Type and Transgenic Arabidopsis Plants. *The Plant Cell*, 15(12), 2866–2884.

Ahmad, S., Scopes, R.K., Rees, G.N., and Patel, B.K.C., 2000. Saccharococcus caldoxylosilyticus sp. nov., an obligately thermophilic, xylose-utilizing, endospore-forming bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 50(2), 517–523.

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., and Stephanopoulos, G., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science*, 330(6000), 70–74.

Alain, K., Marteinsson, V.T., Miroshnichenko, M.L., Bonch-Osmolovskaya, E.A., Prieur, D., and Birrien, J.L., 2002. Marinitoga piezophila sp. nov., a rod-shaped, thermo-piezophilic bacterium isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology*, 52(4), 1331–1339.

Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J., and Mortensen, U.H., 2011. Diversion of flux toward sesquiterpene production in Saccharomyces cerevisiae by fusion of host and heterologous enzymes. *Applied and Environmental Microbiology*, 77(3), 1033–40.

Albuquerque, L., Rainey, F.A., Chung, A.P., Sunna, A., Nobre, M.F., Grote, R., Antranikian, G., and Da Costa, M.S., 2000. Alicyclobacillus hesperidum sp. nov. and a related genomic species from solfataric soils of São Miguel in the Azores. *International Journal of Systematic and Evolutionary Microbiology*, 50(2), 451–457.

Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M. V, Renfrew, P.D., Mulligan, V.K., Kappel, K., Labonte, J.W., Pacella, M.S., Bonneau, R., Bradley, P., Dunbrack Jr., R.L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J.J., 2017. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13, 3031–3048.

Alonso-Gutierrez, J., Chan, R., Batth, T.S., Adams, P.D., Keasling, J.D., Petzold, C.J., and Lee, T.S., 2013. Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production. *Metabolic Engineering*, 19, 33–41.

Alonso-Gutierrez, J., Kim, E.-M., Batth, T.S., Cho, N., Hu, Q., Chan, L.J.G., Petzold, C.J., Hillson, N.J., Adams, P.D., Keasling, J.D., Martin, H.G., and Lee, T.S., 2015. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metabolic Engineering*, 28, 123–133.

Amiri, P., Shahpiri, A., Asadollahi, M.A., Momenbeik, F., and Partow, S., 2016. Metabolic engineering of Saccharomyces cerevisiae for linalool production. *Biotechnology Letters*, 38(3), 503–508.

Anbar, M., Gul, O., Lamed, R., Sezerman, U.O., and Bayer, E.A., 2012. Improved Thermostability of Clostridium thermocellum Endoglucanase Cel8A by Using Consensus-Guided Mutagenesis. *Applied and Environmental Microbiology*, 78(9), 3458–3464.

Anbar, M., Lamed, R., and Bayer, E.A., 2010. Thermostability Enhancement of Clostridium thermocellum Cellulosomal Endoglucanase Cel8A by a Single Glycine Substitution. *ChemCatChem*, 2(8), 997–1003.

Arendt, P., Pollier, J., Callewaert, N., and Goossens, A., 2016. Synthetic biology for production of natural and new-to-nature terpenoids in photosynthetic organisms. *The Plant Journal*, 87(1), 16–37.

Arfman, N., Dijkhuizen, L., Kirchhof, G., Ludwig, W., Schleifer, K.-H., Bulygina, E.S., Chumakov, K.M., Govorukhina, N.I., Trotsenko, Y.A., White, D., and Sharp, R.J., 1992. Bacillus methanolicus sp. nov., a New Species of Thermotolerant, Methanol-Utilizing, Endospore-Forming Bacteria. *International Journal of Systematic Microbiology*, 42(3), 439–445.

Azami, Y., Hattori, A., Nishimura, H., Kawaide, H., Yoshimura, T., and Hemmi, H., 2014. (R)-mevalonate 3-phosphate is an intermediate of the mevalonate pathway in Thermoplasma acidophilum. *The Journal of Biological Chemistry*, 289(23), 15957–67.

de Azevedo, S.M., Faria, M.V., Maluf, W.R., de Oliveira, A.C.B., and de Freitas, J.A., 2003. Zingiberene-mediated resistance to the South American tomato pinworm derived from Lycopersicon hirsutum var. hirsutum. *Euphytica*, 134(3), 347–351.

Azhar, S.H.M., Abdulla, R., Jambo, S.A., Marbawi, H., Gansau, J.A., Faik, A.A.M., and Rodrigues, K.F., 2017. Yeasts in sustainable bioethanol production: A review. *Biochemistry and Biophysics Reports*, 10, 52–61.

Badieyan, S., Bevan, D.R., and Zhang, C., 2012. Study and Design of Stability in GH5 Cellulases. *Biotechnology and Bioengineering*, 109(1), 31–44.

Baer, P., Rabe, P., Citron, C.A., de Oliveira Mann, C.C., Kaufmann, N., Groll, M., and Dickschat, J.S., 2014. Hedycaryol synthase in complex with nerolidol reveals terpene cyclase mechanism. *ChemBioChem*, 15(2), 213–216.

Baer, P., Rabe, P., Fischer, K., Citron, C.A., Klapschinski, T.A., Groll, M., and Dickschat, J.S., 2014. Induced-Fit Mechanism in class I Terpene Cyclases. *Angewandte Chemie International Edition*, 53(29), 7652–7656.

Bakkali, F., Averbeck, S., Averbeck, D., and Idaomar, M., 2008. Biological effects of essential oils - A review. *Food and Chemical Toxicology*, 46(2), 446–475.

Balibar, C.J., Shen, X., and Tao, J., 2009. The mevalonate pathway of

Staphylococcus aureus. *Journal of bacteriology*, 191(3), 851–61.

Banta, A.B., Wei, J.H., Gill, C.C.C., Giner, J.-L., and Welander, P. V, 2017. Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *PNAS*, 114(2), 245–250.

Beale, M.H., Birkett, M.A., Bruce, T.J.A., Chamberlain, K., Field, L.M., Huttly, A.K., Martin, J.L., Parker, R., Phillips, A.L., Pickett, J.A., Prosser, I.M., Shewry, P.R., Smart, L.E., Wadhams, L.J., Woodcock, C.M., and Zhang, Y., 2006. Aphid alarm pheromone produced by transgenic plants affects aphid and parasitoid behavior. *PNAS*, 103(27), 10509–10513.

Beeder, J., Torsvik, T., and Lien, T., 1995. Thermodesulforhabdus norvegicus gen. nov., sp. nov., a novel thermophilic sulfate-reducing bacterium from oil field water. *Archives of Microbiology*, 164, 331–336.

van Beek, T.A. and Joulain, D., 2018. The essential oil of patchouli, Pogostemon cablin: A review. *Flavour and Fragrance Journal*, 33(1), 6–51.

Beekwilder, J., van Houwelingen, A., Cankar, K., van Dijk, A.D.J., de Jong, R.M., Stoopen, G., Bouwmeester, H., Achkar, J., Sonke, T., and Bosch, D., 2014. Valencene synthase from the heartwood of Nootka cypress (Callitropsis nootkatensis) for biotechnological production of valencene. *Plant Biotechnology Journal*, 12(2), 174–182.

Beller, H.R., Lee, T.S., and Katz, L., 2015. Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Natural Product Reports*, 32(10), 1508–1526.

Bentley, F.K. and Melis, A., 2012. Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms. *Biotechnology and Bioengineering*, 109(1), 100–109.

Beran, F., Rahfeld, P., Luck, K., Nagel, R., Vogel, H., Wielsch, N., Irmisch, S., Ramasamy, S., Gershenzon, J., Heckel, D.G., and Köllner, T.G., 2016. Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. *PNAS*, 113(11), 2922–2927.

Bezuidt, O.K.I., Gomri, M.A., Pierneef, R., van Goethem, M.W., Kharroub, K., Cowan, D.A., and Makhalanyane, T.P., 2016. Draft genome sequence of Thermoactinomyces sp. strain AS95 isolated from a Sebkha in Thamelaht,

Algeria. *Standards in Genomic Sciences*, 11(1), 1–6.

Bhatia, S.P., McGinty, D., Letizia, C.S., and Api, A.M., 2008. Fragrance material review on sclareol. *Food and Chemical Toxicology*, 46, S270–S274.

Bian, G., Han, Y., Hou, A., Yuan, Y., Liu, X., Deng, Z., and Liu, T., 2017. Releasing the potential power of terpene synthases by a robust precursor supply platform. *Metabolic Engineering*, 42, 1–8.

Blank, P.N., Barrow, G.H., Chou, W.K.W., Duan, L., Cane, D.E., and Christianson, D.W., 2017. Substitution of Aromatic Residues with Polar Residues in the Active Site Pocket of epi-Isozizaene Synthase Leads to the Generation of New Cyclic Sesquiterpenes. *Biochemistry*, 56(43), 5798–5811.

Bleeker, P.M., Diergaarde, P.J., Ament, K., Schütz, S., Johne, B., Dijkink, J., Hiemstra, H., de Gelder, R., de Both, M.T.J., Sabelis, M.W., Haring, M.A., and Schuurink, R.C., 2010. Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. *Phytochemistry*, 72(1), 68–73.

Blerot, B., Martinelli, L., Prunier, C., Saint-Marcoux, D., Legrand, S., Bony, A., Sarrabère, L., Gros, F., Boyer, N., Caissard, J.-C., Baudino, S., and Jullien, F., 2018. Functional Analysis of Four Terpene Synthases in Rose-Scented Pelargonium Cultivars (Pelargonium × hybridum) and Evolution of Scent in the Pelargonium Genus. *Frontiers in Plant Science*, 9, 1–18.

Bloudoff, K. and Schmeing, T.M., 2017. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1865(11), 1587–1604.

de Boer, L., Dijkhuizen, L., Grobben, G., Goodfellow, M., Stackebrandt, E., Parlett, J.H., Whitehead, D., and Witt, D., 1990. Amycolatopsis methanolica sp. nov., a Facultatively Methylotrophic Actinomycete. *International Journal of Systematic Bacteriology*, 40(2), 194–204.

Bogdanova, T.I., Tsaplina, I.A., Kondrat'eva, T.F., Duda, V.I., Suzina, N.E., Melamud, V.S., Tourova, T.P., and Karavaiko, G.I., 2006. Sulfobacillus thermotolerans sp. nov., a thermotolerant chemolithotrophic bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 56(5), 1039–1042.

Bohlmann, J., Meyer-Gauen, G., and Croteau, R., 1998. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *PNAS*, 95(8), 4126–4133.

Bohlmann, J., Steele, C.L., and Croteau, R., 1997. Monoterpene Synthases from Grand Fir (Abies grandis). *The Journal of Biological Chemistry*, 272(35), 21784–21792.

Boronat, A. and Rodríguez-Concepción, M., 2014. Terpenoid Biosynthesis in Prokaryotes. *In*: *Biotechnology of Isoprenoids*. 3–18.

Bosma, E.F., van de Weijer, A.H.P., Daas, M.J.A., van der Oost, J., de Vos, W.M., and van Kranenburg, R., 2015. Isolation and Screening of Thermophilic Bacilli from Compost for Electrotransformation and Fermentation: Characterization of *Bacillus smithii* ET 138 as a New Biocatalyst. *Applied and Environmental Microbiology*, 81(5), 1874–1883.

Bottini, R., Cassán, F., and Piccoli, P., 2004. Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. *Applied Microbiology and Biotechnology*, 65(5), 497–503.

Bowers, W.S., Nault, L.R., Webb, R.E., and Dutky, S.R., 1972. Aphid Alarm Pheromone: Isolation, Identification, Synthesis. *Science*, 177(4054), 1121–1122.

Brandle, J.E. and Telmer, P.G., 2007. Steviol glycoside biosynthesis. *Phytochemistry*, 68(14), 1855–1863.

Brennan, T.C.R., Turner, C.D., Krömer, J.O., and Nielsen, L.K., 2012. Alleviating monoterpene toxicity using a two-phase extractive fermentation for the bioproduction of jet fuel mixtures in Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, 109(10), 2513–2522.

Bringi, V., Kadkade, P.G., Kane, E.J., Prince, C.L., Roach, B., and Schubmehl, B.F., 2003. Enhanced production of taxol and taxanes by cell cultures of taxus species. *Patent EP 0960944 B1*.

Brooks, B.R., Iii, C.L.B., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M., and Karplus, M., 2009. CHARMM : The Biomolecular Simulation Program. *Journal of Computational Chemistry*, 30(10), 1545–1614.

Brown, I., Allen, C., Mummey, D., Sarkisova, S., and Mckay, D., 2007. Iron-Tolerant

Cyanobacteria: Implications for Astrobiology. *In*: *Algae and Cyanobacteria in Extreme Environments*. 425–442.

Brumm, P., Land, M.L., Hauser, L.J., Jeffries, C.D., Chang, Y.-J., and Mead, D.A., 2015. Complete genome sequences of Geobacillus sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, 10(81).

Buckley, S.A., Clark, K.A., and Evershed, R.P., 2004. Complex Organic Chemical Balms of Pharaonic Animal Mummies. *Nature*, 431(7006), 294–299.

Calhoun, J.R., Kono, H., Lahr, S., Wang, W., Degrado, W.F., and Saven, J.G., 2003. Computational Design and Characterization of a Monomeric Helical Dinuclear Metalloprotein. *Journal of Molecular Biology*, 334, 1101–1115.

Cane, D.E., Oliver, J.S., Harrison, P.H.M., Abell, C., Hubbard, B.R., Kane, C.T., and Lattman, R., 1990. Biosynthesis of Pentalenene and Pentalenolactone. *Journal of the American Chemical Society*, 112(11), 4513–4524.

Cane, D.E., Shim, J.H., Xue, Q., Fitzsimons, B.C., and Hohn, T.M., 1995. Trichodiene Synthase. Identification of Active Site Residues by Site-Directed Mutagenesist. *Biochemistry*, 34(8), 2480–2488.

Cane, D.E., Sohng, J.-K., Lamberson, C.R., Rudnicki, S.M., Wu, Z., Lloyd, M.D., Oliver, J.S., and Hubbard, B.R., 1994. Pentalenene Synthase. Purification, Molecular Cloning, Sequencing, and High-Level Expression in Escherichia coli of a Terpenoid Cyclase from Streptomyces UC53191. *Biochemistry*, 33, 5846–5857.

Caniard, A., Zerbe, P., Legrand, S., Cohade, A., Valot, N., Magnard, J.-L., Bohlmann, J., and Legendre, L., 2012. Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in Salvia sclarea (L.) and their relevance for perfume manufacture. *BMC Plant Biology*, 12, 1–13.

Cao, R., Zhang, Y., Mann, F.M., Huang, C., Mukkamala, D., Hudock, M.P., Mead, M.E., Prisic, S., Wang, K., Lin, F.-Y.Y., Chang, T.-K.K., Peters, R.J., and Oldfield, E., 2010. Diterpene cyclases and the nature of the isoprene fold. *Proteins: Structure, Function and Bioinformatics*, 78(11), 2417–2432.

Capriotti, E., Fariselli, P., Calabrese, R., and Casadio, R., 2005. Protein Structure and Function Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, 21, 54–58.

Caputi, L. and Aprea, E., 2011. Use of terpenoids as natural flavouring compounds in food industry. *Recent patents on food, nutrition & agriculture*, 3(1), 9–16.

Caron, B., Mark, A.E., and Poger, D., 2014. Some Like It Hot: The Effect of Sterols and Hopanoids on Lipid Ordering at High Temperature. *The Journal of Physical Chemistry Letters*, 5(22), 3953–3957.

Carreto, L., Moore, E., Nobre, F.M., Wait, R., Riley, P.W., Sharp, R.J., and Da Costa, M.S., 1996. Rubrobacter xylanophilus sp. nov., a New Thermophilic Species Isolated from a Thermally Polluted Effluent. *International Journal of Systematic Bacteriology*, 46(2), 460–465.

Case, D.A., Cheatham, T.E., Darden, T.O.M., Gohlke, H., Luo, R.A.Y., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J., 2005. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*, 26(16), 1668–1688.

Catici, D.A.M., Horne, J.E., Cooper, G.E., and Pudney, C.R., 2015. Polyubiquitin Drives the Molecular Interactions of the NF-κB Essential Modulator (NEMO) by Allosteric Regulation. *The Journal of Biological Chemistry*, 290(22), 14130–9.

Chacón, M.G., Marriott, A., Kendrick, E.G., Styles, M.Q., and Leak, D.J., 2019. Esterification of geraniol as a strategy for increasing product titre and specificity in engineered Escherichia coli. *Microbial Cell Factories*, 18, 1–11.

Chan Joo, J., Pack, S.P., Kim, Y.H., and Yoo, Y.J., 2011. Thermostabilization of Bacillus circulans xylanase: Computational optimization of unstable residues based on thermal fluctuation analysis. *Journal of Biotechnology*, 151, 56–65.

Chandran, S.S., Kealey, J.T., and Reeves, C.D., 2011. Microbial production of isoprenoids. *Process Biochemistry*, 46, 1703–1710.

Chang, M.C.Y. and Keasling, J.D., 2006. Production of isoprenoid pharmaceuticals by engineered microbes. *Nature Chemical Biology*, 2, 674–681.

Chapman, J., Ismail, A.E., and Dinu, C.Z., 2018. Industrial applications of enzymes: Recent advances, techniques, and outlooks. *Catalysts*, 8(6), 1–26.

Chen, F., Al-Ahmad, H., Joyce, B., Zhao, N., Köllner, T.G., Degenhardt, J., and Stewart, C.N., 2009. Within-plant distribution and emission of sesquiterpenes from Copaifera officinalis. *Plant Physiology and Biochemistry*, 47(11–12), 1017–1023.

Chen, J.-J., Lin, L.-B., Zhang, L.-L., Zhang, J., Tang, S.-K., Wei, Y.-L., and Li, W.-J., 2012. Laceyella sediminis sp. nov., a thermophilic bacterium isolated from a hot spring. *International Journal of Systematic and Evolutionary Microbiology*, 62, 38–42.

Chen, J., Yu, H., Liu, C., Liu, J., and Shen, Z., 2012. Improving stability of nitrile hydratase by bridging the salt-bridges in specific thermal-sensitive regions. *Journal of Biotechnology*, 164, 354–362.

Chen, W. and Viljoen, A.M., 2010. Geraniol-A review of a commercially important fragrance material. *South African Journal of Botany*, 76(4), 643–651.

Chen, X.-Y., Chen, Y., Heinstein, P., and Davisson, V.J., 1995. Cloning, Expression, and Characterization of (+)-δ-Cadinene Synthase: A Catalyst for Cotton Phytoalexin Biosynthesis. *Archives of Biochemistry and Biophysics*, 324(2), 255–266.

Chen, Y., He, Y., Zhang, B., Yang, J., Li, W., Dong, Z., and Hu, S., 2011. Complete Genome Sequence of Alicyclobacillus acidocaldarius Strain Tc-4-1. *Journal of Bacteriology*, 193(19), 5602–5603.

Cheng, F. and Cheng, Z., 2015. Research Progress on the use of Plant Allelopathy in Agriculture and the Physiological and Ecological Mechanisms of Allelopathy. *Frontiers in Plant Science*, 6, 1–16.

Chertkov, O., Sikorski, J., Nolan, M., Lapidus, A., Lucas, S., Glavina, T., Rio, D., Tice, H., Cheng, J.-F., Goodwin, L., Pitluck, S., Liolios, K., Ivanova, N., Mavromatis, K., Mikhailova, N., Ovchinnikova, G., Pati, A., Chen, A., Palaniappan, K., Djao, O.D.N., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Brettin, T., Han, C., Detter, J.C., Rohde, M., Göker, M., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., and Kyrpides, N.C., 2011. Complete genome sequence of Thermomonospora curvata type strain (B9T). *Standards in Genomic Sciences*, 4(1), 13–22.

Chiba, R., Minami, A., Gomi, K., and Oikawa, H., 2013. Identification of Ophiobolin F Synthase by a Genome Mining Approach: A Sesterterpene Synthase from Aspergillus clavatus. *Organic Letters*, 15(3), 594–597.

Chomel, M., Guittonny-Larchevêque, M., Fernandez, C., Gallet, C., DesRochers, A., Paré, D., Jackson, B.G., and Baldy, V., 2016. Plant secondary metabolites: a key driver of litter decomposition and soil nutrient cycling. *Journal of Ecology*,

104, 1527–1541.

Chotani, G.K., McAuliffe, J.C., Miller, M.C., Muir, R.E., Vaviline, D. V, and Weyler, W., 2013. Isoprene Production using the DXP and MVA Pathway. *Patent US 8507235 B2*.

Chou, W.K.W., Fanizza, I., Uchiyama, T., Komatsu, M., Ikeda, H., and Cane, D.E., 2010. Genome mining in Streptomyces avermitilis: Cloning and characterization of SAV_76, the synthase for a new sesquiterpene, avermitilol. *Journal of the American Chemical Society*, 132(26), 8850–8851.

Chouhan, S., Sharma, K., and Guleria, S., 2017. Antimicrobial Activity of Some Essential Oils-Present Status and Future Perspectives. *Medicines (Basal)*, 4(3), 1–21.

Christianson, D.W., 2007. Roots of Biosynthetic Diversity. *Science*, 316(5821), 60–61.

Christianson, D.W., 2008. Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology*, 12(2), 141–150.

Christianson, D.W., 2017. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*, 117(17), 11570–11648.

Citron, C.A., Gleitzmann, J., Laurenzano, G., Pukall, R., and Dickschat, J.S., 2012. Terpenoids are Widespread in Actinomycetes: A Correlation of Secondary Metabolism and Genome Data. *ChemBioChem*, 13(2), 202–214.

Cline, J., Braman, J.C., and Hogrefe, H.H., 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24(18), 3546–3551.

Cobb, R.E., Chao, R., and Zhao, H., 2013. Directed Evolution: Past, Present and Future. *AIChE Journal*, 59(5), 1432–1440.

Cole, J.K., Gieler, B.A., Heisler, D.L., Palisoc, M.M., Williams, A.J., Dohnalkova, A.C., Ming, H., Yu, T.T., Dodsworth, J.A., Li, W.-J., and Hedlund, B.P., 2013. Kallotenue papyrolyticum gen. nov., sp. nov., a cellulytic and filamentous thermophile that represents a novel lineage (Kallotenuales ord. nov., Kallotenuaceae fam. nov.) within the class Chloroflexia. *International Journal of Systematic and Evolutionary Microbiology*, 63(12), 4675–4682.

Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin,

S., and Atkinson, T., 2009. Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production. *Metabolic Engineering*, 11(6), 398–408.

Crock, J., Wildung, M., and Croteau, R., 1997. Isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (Mentha x piperita, L.) that produces the aphid alarm pheromone (E)-β-farnesene. *Biochemistry*, 94(24), 12833–12838.

Croteau, R. and Cane, D.E., 1985. Monoterpene and Sesquiterpene Cyclases. *In*: *Methods in Enzymology*. 383–405.

Croteau, R., Ketchum, R.E.B., Long, R.M., Kaspera, R., and Wildung, M.R., 2006. Taxol biosynthesis and molecular genetics. *Phytochemistry Reviews*, 5(1), 75–97.

Crowell, A.L., Williams, D.C., Davis, E.M., Wildung, M.R., and Croteau, R., 2002. Molecular cloning and characterization of a new linalool synthase. *Archives of Biochemistry and Biophysics*, 405(1), 112–121.

Cushnie, T.P.T., Cushnie, B., and Lamb, A.J., 2014. Alkaloids: An overview of their antibacterial, antibiotic-enhancing and antivirulence activities. *International Journal of Antimicrobial Agents*, 44(5), 377–386.

D'este, M., Alvarado-Morales, M., and Angelidaki, I., 2018. Amino acids production focusing on fermentation technologies - A review. *Biotechnology Advances*, 36(1), 14–25.

Dairi, T., Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., and Seto, H., 2001. Eubacterial Diterpene Cyclase Genes Essential for Production of the Isoprenoid Antibiotic Terpentecin. *Journal of Bacteriology*, 183(20), 6085–6094.

Dang, T. and Prestwich, G.D., 2000. Site-directed mutagenesis of squalene-hopene cyclase: altered substrate specificity and product distribution. *Chemistry & Biology*, 7(8), 643–649.

Deguerry, F., Pastore, L., Wu, S., Clark, A., Chappell, J., and Schalk, M., 2006. The diverse sesquiterpene profile of patchouli, Pogosemon cablin, is correlated with a limited number of sesquiterpene synthases. *Archives of Biochemistry and Biophysics*, 454(2), 123–136.

Deinhard, G., Blanz, P., Poralla, K., and Altan, E., 1987. Bacillus acidoterrestris sp.

nov., a New Thermotolerant Acidophile Isolated from Different Soils. *Systematic and Applied Microbiology*, 10(1), 47–53.

Deligeorgopoulou, A. and Allemann, R.K., 2003. Evidence for Differential Folding of Farnesyl Pyrophosphate in the Active Site of Aristolochene Synthase: A Single-Point Mutation Converts Aristolochene Synthase into an (E)-β-Farnesene Synthase. *Biochemistry*, 42(25), 7741–7747.

Dellas, N., Thomas, S.T., Manning, G., and Noel, J.P., 2013. Discovery of a metabolic alternative to the classical mevalonate pathway. *eLife*, 2, 1–18.

DeLuca, S., Khar, K., and Meiler, J., 2015. Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS ONE*, 10(7), 1–19.

Demain, A.L., 2007. The business of biotechnology. *Industrial Biotechnology*, 3(3), 269–283.

Demain, A.L. and Vaishnav, P., 2010. Natural products for cancer chemotherapy. *Microbial Biotechnology*, 4(6), 687–699.

Demissie, Z.A., Tarnowycz, M., Adal, A.M., Sarker, L.S., and Mahmoud, S.S., 2019. A lavender ABC transporter confers resistance to monoterpene toxicity in yeast. *Planta*, 249(1), 139–144.

Despinasse, Y., Fiorucci, S., Antonczak, S., Moja, S., Bony, A., Nicolè, F., Baudino, S., Magnard, J.-L., and Jullien, F., 2017. Bornyl-diphosphate synthase from Lavandula angustifolia: A major monoterpene synthase involved in essential oil quality. *Phytochemistry*, 137, 24–33.

Diamond, A. and Desgagné-Penix, I., 2016. Metabolic engineering for the production of plant isoquinoline alkaloids. *Plant Biotechnology Journal*, 14(6), 1319–1328.

Diaz, J.E., Lin, C.S., Kunishiro, K., Feld, B.K., Avrantinis, S.K., Bronson, J., Greaves, J., Saven, J.G., and Weiss, G.A., 2011. Computational design and selections for an engineered, thermostable terpene synthase. *Protein Science*, 20(9), 1597–1606.

Dickschat, J.S., 2016. Bacterial terpene cyclases. *Natural Product Reports*, 33(1), 87–110.

Dickschat, J.S., Pahirulzaman, K.A.K., Rabe, P., and Klapschinski, T.A., 2014. An Improved Technique for the Rapid Chemical Characterisation of Bacterial Terpene Cyclases. *ChemBioChem*, 15(6), 810–814.

Ding, M., Yan, H., Li, L., Zhai, F., Shang, L., Yin, Z., and Yuan, Y., 2014. Biosynthesis of Taxadiene in Saccharomyces cerevisiae: Selection of Geranylgeranyl Diphosphate Synthase Directed by a Computer-Aided Docking Strategy. *PLoS ONE*, 9(10), 1–9.

Dinsdale, A.E., Halket, G., Coorevits, A., Landschoot, A. Van, Busse, H.-J., de Vos, P., and Logan, N.A., 2011. Emended descriptions of Geobacillus thermoleovorans and Geobacillus thermocatenulatus. *International Journal of Systematic and Evolutionary Microbiology*, 61(8), 1802–1810.

Doi, K., Mori, K., Martono, H., Nagayoshi, Y., Fujino, Y., Tashiro, K., Kuhara, S., and Ohshima, T., 2013. Draft Genome Sequence of Geobacillus kaustophilus GBlys, a Lysogenic Strain with Bacteriophage ØOH2. *Genome Announcements*, 1(4).

Du, F.-L., Yu, H.-L., Xu, J.-H., and Li, C.-X., 2014. Enhanced limonene production by optimizing the expression of limonene biosynthesis and MEP pathway genes in E. coli. *Bioresources and Bioprocessing*, 1, 1–10.

Duell, E.R., D'Agostino, P.M., Shapiro, N., Woyke, T., Fuchs, T.M., and Gulder, T.A.M., 2019. Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in E. coli. *Microbial Cell Factories*, 18, 1–11.

Durand, H.A.G. and Durand, F., 2009. Plant Extract Obtained by an Extraction Method by Means of Solvents of Plant Origin. *Patent US 20090306386 A1*.

Dürr, C., Schnell, H.-J., Luzhetskyy, A., Murillo, R., Weber, M., Welzel, K., Vente, A., and Bechthold, A., 2006. Biosynthesis of the Terpene Phenalinolactone in Streptomyces sp. Tü 6071: Analysis of the Gene Cluster and Generation of Derivatives. *Chemistry & Biology*, 13, 365–377.

Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., and Finn, R.D., 2018. The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, 427–432.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S., 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE*, 4(5), e5553.

Engler, C., Kandzia, R., and Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE*, 3(11), e3647.

Fähnrich, A., Neumann, M., and Piechulla, B., 2014. Characteristic alatoid 'cineole cassette' monoterpene synthase present in Nicotiana noctiflora. *Plant Molecular Biology*, 85(2), 135–145.

Fardeau, M.-L., Bonilla Salinas, M., L'Haridon, S., Jeanthon, C., Verhé, F., Cayol, J.L., Patel, B.K.C., Garcia, J.L., and Ollivier, B., 2004. Isolation from oil resorvoirs of novel thermophilic anaerobes phylogenetically related to Thermoanaerobacter subterraneus. *International Journal of Systematic and Evolutionary Microbiology*, 54(2), 467–474.

Feil, C., Süssmuth, R., Jung, G., and Poralla, K., 1996. Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. *European Journal of Biochemistry*, 242, 51–55.

Felicetti, B. and Cane, D.E., 2004. Aristolochene Synthase: Mechanistic Analysis of Active Site Residues by Site-Directed Mutagenesis. *Journal of American Chemical Society*, 126, 7212–7221.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R., and Wang, L., 2007. Genome and proteome of long-chain alkane degrading Geobacillus thermodenitrificans NG80-2 isolated from a deep-subsurface oil reservoir. *PNAS*, 104(13), 5602–5607.

Feyhl-Buska, J., Chen, Y., Jia, C., Wang, J.-X., Zhang, C.L., and Boyd, E.S., 2016. Influence of Growth Phase, pH, and Temperature on the Abundance and Composition of Tetraether Lipids in the Thermoacidophile *Picrophilus torridus*. *Frontiers in Microbiology*, 7, 1–12.

Filippidou, S., Jaussi, M., Junier, T., Wunderlin, T., Jeanneret, N., Palmieri, F., Palmieri, I., Roussel-Delif, L., Vieth-Hillebrand, A., Vetter, A., Chain, P.S., Regenspurg, S., and Junier, P., 2016. Anoxybacillus geothermalis sp. nov., a facultatively anaerobic, endospore-forming bacterium isolated from mineral deposits in a geothermal station. *International Journal of Systematic and Evolutionary Microbiology*, 66(8), 2944–2951.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R., 2015. HMMER web server: 2015 update. *Nucleic Acids Research*, 43(W1), W30–W38.

Flores-Sanchez, I.J. and Verpoorte, R., 2009. Plant Polyketide Synthases: A fascinating group of enzymes. *Plant Physiology and Biochemistry*, 47(3), 167–174.

Francis, F., Lognay, G., and Haubruge, E., 2004. OLFACTORY RESPONSES TO APHID AND HOST PLANT VOLATILE RELEASES: (E)-β-FARNESENE AN EFFECTIVE KAIROMONE FOR THE PREDATOR Adalia bipunctata. *Journal of Chemical Ecology*, 30(4), 741–755.

Francis, F., Vandermoten, S., Verheggen, F., Lognay, G., and Haubruge, E., 2005. Is the (E)-β-farnesene only volatile terpenoid in aphids? *Journal of Applied Entomology*, 129(1), 6–11.

Fraser, M.D., Davies, J.R.T., and Chang, X., 2017. New Gold in Them Thar Hills: Testing a Novel Supply Route for Plant-Derived Galanthamine. *Journal of Alzheimer's Disease*, 55, 1321–1325.

Frickey, T. and Kannenberg, E., 2009. Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. *Environmental Microbiology*, 11(5), 1224–1241.

Galzitskaya, O. V, Garbuzynskiy, S.O., and Lobanov, M.Y., 2006. FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics*, 22(23), 2948–2949.

Gao, Y., Honzatko, R.B., and Peters, R.J., 2013. Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Natural Product Reports*, 29(10), 1153–1175.

Garcia Costas, A.M., Tsukatani, Y., Rijpstra, W.I.C., Schouten, S., Welander, P. V, Summons, R.E., and Bryant, D.A., 2012. Identification of the Bacteriochlorophylls, Carotenoids, Quinones, Lipids, and Hopanoids of 'Candidatus Chloracidobacterium thermophilum'. *Journal of Bacteriology*, 194(5), 1158–1168.

Garrett, S.R., Morris, R.J., and O'Maille, P.E., 2012. Steady-State Kinetic Characterization of Sesquiterpene Synthases by Gas Chromatography-Mass Spectroscopy. *In*: *Methods in Enzymology*. Elsevier Inc., 3–19.

Gatson, J.W., Benz, B.F., Chandrasekaran, C., Satomi, M., Venkateswaran, K., and Hart, M.E., 2006. Bacillus tequilensis sp. nov., isolated from a 2000-year-old Mexican shaft-tomb, is closely related to Bacillus subtilis. *International Journal*

of Systematic and Evolutionary Microbiology, 56(7), 1475–1484.

Gerber, N.N. and Lechevalier, H.A., 1965. Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes. *Applied Microbiology*, 13(6), 935–938.

Gibson, D.G., Young, L., Chuang, R., Venter, J.C., Hutchison III, C.A., and Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345.

Giglio, S., Chou, W.K.W., Ikeda, H., Cane, D.E., and Monis, P.T., 2011. Biosynthesis of 2-Methylisoborneol in Cyanobacteria. *Environmental Science & Technology*, 45(3), 992–998.

Glyakina, A. V, Garbuzynskiy, S.O., Lobanov, M.Y., and Galzitskaya, O. V, 2007. Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. *Bioinformatics*, 23(17), 2231–2238.

Goldsmith, N., Hansen, E.H., Meyer, J.-P., and Brianza, F., 2015. Process for Producing Vanillin. *Patent WO 2015121379 A3*.

Gong, H.-Y., Zeng, Y., and Chen, X.-Y., 2014. Diterpene Synthases and Their Responsible Cyclic Natural Products. *Natural Products and Bioprospecting*, 4(2), 59–72.

Gray, D.W., Breneman, S.R., Topper, L.A., and Sharkey, T.D., 2011. Biochemical Characterization and Homology Modeling of Methylbutenol Synthase and Implications for Understanding Hemiterpene Synthase Evolution in Plants. *The Journal of Biological Chemistry*, 286(23), 20582–20590.

Gray, D.W., Goldstein, A.H., and Lerdau, M.T., 2006. Thermal history regulates methylbutenol basal emission rate in Pinus ponderosa. *Plant, Cell and Environment*, 29(7), 1298–1308.

Greenfield, N.J., 2007. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1(6), 2876–2890.

Greenhagen, B.T., O'Maille, P.E., Noel, J.P., and Chappell, J., 2006. Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *PNAS*, 103(26), 9826–9831.

de Groot, N.S. and Ventura, S., 2006. Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Letters*.

Grundy, D.J., Chen, M., González, V., Leoni, S., Miller, D.J., Christianson, D.W., and Allemann, R.K., 2016. Mechanism of Germacradien-4-ol Synthase-Controlled Water Capture. *Biochemistry*, 55(14), 2112–2121.

Guerois, R., Nielsen, J.E., and Serrano, L., 2002. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *Journal of Molecular Biology*, 320, 369–387.

Gumulya, Y. and Reetz, M.T., 2011. Enhancing the Thermal Robustness of an Enzyme by Directed Evolution: Least Favorable Starting Points and Inferior Mutants Can Map Superior Evolutionary Pathways. *ChemBioChem*, 12(16), 2502–2510.

Guo, R.-T., Kuo, C.-J., Chou, C.-C., Ko, T.-P., Shr, H.-L., Liang, P.-H., and Wang, A.H.-J., 2004. Crystal Structure of Octaprenyl Pyrophosphate Synthase from Hyperthermophilic Thermotoga maritima and Mechanism of Product Chain Length Determination. *The Journal of Biological Chemistry*, 279(6), 4903–4912.

Guo, R.-T., Kuo, C.-J., Ko, T.-P., Chou, C.-C., Liang, P.-H., and Wang, A.H.-J., 2004. A Molecular Ruler for Chain Elongation Catalyzed by Octaprenyl Pyrophosphate Synthase and Its Structure-Based Engineering To Produce Unprecedented Long Chain *trans*-Prenyl Products. *Biochemistry*, 43(24), 7678–7686.

Gupta, P. and Phulara, S.C., 2015. Metabolic engineering for isoprenoid-based biofuel production. *Journal of Applied Microbiology*, 119(3), 605–619.

Gurung, N., Ray, S., Bose, S., and Rai, V., 2013. A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. *BioMed Research International*, 2013.

Hale, V., Keasling, J.D., Renninger, N., and Diagana, T.T., 2007. Microbially Derived Artemisinin: A Biotechnology Solution to the Global Problem of Access to Affordable Antimalarial Drugs. *American Journal of Tropical Medicine and Hygiene*, 77(6), 198–202.

Halpin, C.M., Reilly, C., and Walsh, J.J., 2010. Nature's Anti-Alzheimer's Drug: Isolation and Structure Elucidation of Galantamine from Leucojum aestivum. *Journal of Chemical Education*, 87(11), 1242–1243.

Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., Seto, H., and Dairi, T., 2002. Functional Analysis of Eubacterial Diterpene Cyclases Responsible for Biosynthesis of a Diterpene Antibiotic, Terpentecin. *The Journal of Biological*

Chemistry, 277(40), 37098–37104.

Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A., and Bohlmann, J., 2011. Evolution of Diterpene Metabolism: Sitka Spruce CYP720B4 Catalyzes Multiple Oxidations in Resin Acid Biosynthesis of Conifer Defense against Insects. *Plant Physiology*, 157(4), 1677–1695.

Hansen, J., 2011. Method Of Producing Isoprenoid Compounds In Yeast. *Patent WO 2011146833 A1*.

Hanson, J.R., 2009. Diterpenoids. *Natural Product Reports*, 26(9), 1156–1171.

Hare, J.D., 2011. Ecological Role of Volatiles Produced by Plants in Response to Damage by Herbivorous Insects. *Annual Review of Entomology*, 56, 161–180.

Harmel, N., Almohamad, R., Fauconnier, M.L., Jardin, P. Du, Verheggen, F., Marlier, M., Haubruge, E., and Francis, F., 2007. Role of terpenes from aphid-infested potato on searching and oviposition behavior of Episyrphus balteatus. *Insect Science*, 14(1), 57–63.

Harris, G.G., Lombardi, P.M., Pemberton, T.A., Matsui, T., Weiss, T.M., Cole, K.E., Köksal, M., Murphy, F. V, Vedula, L.S., Chou, W.K.W., Cane, D.E., and Christianson, D.W., 2015. Structural Studies of Geosmin Synthase, a Bifunctional Sesquiterpene Synthase with αα Domain Architecture That Catalyzes a Unique Cyclization–Fragmentation Reaction Sequence. *Biochemistry*, 54(48), 7142–7155.

Harrison, K.W. and Harvey, B.G., 2017. Renewable high density fuels containing tricyclic sesquiterpanes and alkyl diamondoids. *Sustainable Energy & Fuels*, 1(3), 467–473.

Hattan, J.-I., Shindo, K., Ito, T., Shibuya, Y., Watanabe, A., Chie, T., Fumina, O., Sasaki, T., Ishii, J., Kondo, A., and Misawa, N., 2016. Identification of a novel hedycaryol synthase gene isolated from Camellia brevistyla flowers and floral scent of Camellia cultivars. *Planta*, 243, 959–972.

Hayashi, Y., Matsuura, N., Toshima, H., Itoh, N., Ishikawa, J., Mikami, Y., and Dairi, T., 2008. Cloning of the Gene Cluster Responsible for the Biosynthesis of Brasilicardin A, a Unique Diterpenoid. *The Journal of Antibiotics*, 61(3), 164–174.

Haynes, R.K., 2006. From Artemisinin to New Artemisinin Antimalarials: Biosynthesis,

Extraction, Old and New Derivatives, Stereochemistry and Medicinal Chemistry Requirements. *Current Topics in Medicinal Chemistry*, 6(5), 509–537.

Heinrich, M. and Teoh, H.L., 2004. Galanthamine from snowdrop-the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. *Journal of Ethnopharmacology*, 92, 147–162.

Herman, A., Tambor, K., and Herman, A., 2016. Linalool Affects the Antimicrobial Efficacy of Essential Oils. *Current Microbiology*, 72(2), 165–172.

Hider, R.C. and Kong, X., 2010. Chemistry and biology of siderophores. *Natural Product Reports*, 27(5), 637–657.

Hirte, M., Meese, N., Mertz, M., Fuchs, M., and Brück, T.B., 2018. Insights Into the Bifunctional Aphidicolan-16-β-ol Synthase Through Rapid Biomolecular Modeling Approaches. *Frontiers in Chemistry*, 6, 1–10.

Hosfield, D.J., Zhang, Y., Dougan, D.R., Broun, A., Tari, L.W., Swanson, R. V, and Finn, J., 2003. Structural Basis for Bisphosphonate-mediated Inhibition of Isoprenoid Biosynthesis. *The Journal of Biological Chemistry*, 279(10), 8526–8529.

Hou, S., Makarova, K.S., Saw, J.H.W., Senin, P., Ly, B. V, Zhou, Z., Ren, Y., Wang, J., Galperin, M.Y., Omelchenko, M. V, Wolf, Y.I., Yutin, N., Koonin, E. V, Stott, M.B., Mountain, B.W., Crowe, M.A., Smirnova, A. V, Dunfield, P.F., Feng, L., Wang, L., and Alam, M., 2008. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum *Verrucomicrobia*. *Biology Direct*, 3, 1–25.

Hu, P., Tom, L., Singh, A., Thomas, B.C., Baker, B.J., Piceno, Y.M., Andersen, G.L., and Banfield, J.F., 2016. Genome-Resolved Metagenomic Analysis Reveals Roles for Candidate Phyla and Other Microbial Community Members in Biogeochemical Transformations in Oil Reservoirs. *mBio*, 7(1).

Hu, Y., Chou, W.K.W., Hopson, R., and Cane, D.E., 2011. Genome Mining in Streptomyces clavuligerus: Expression and Biochemical Characterization of Two New Cryptic Sesquiterpene Synthases. *Chemistry & Biology*, 18(1), 32–37.

Huang, M., Sanchez-Moreiras, A.M., Abel, C., Sohrabi, R., Lee, S., Gershenzon, J., and Tholl, D., 2012. The major volatile organic compound emitted from Arabidopsis thaliana flowers, the sesquiterpene (E)-β-caryophyllene, is a defense against a bacterial pathogen. *New Phytologist*, 193(4), 997–1008.

Huembelin, M., Beekwilder, M.J., and Kierkels, J.G.T., 2014. Rhodobacter for preparing terpenoids. *Patent WO 2014014339 A2*.

Hussain, A., Ogawa, T., Saito, M., Sekine, T., Nameki, M., Matsushita, Y., Hayashi, T., and Katayama, Y., 2013. Cloning and expression of a gene encoding a novel thermostable thiocyanate-degrading enzyme from a mesophilic alphaproteobacteria strain THI201. *Microbiology*, 159(11), 2294–2302.

Hussein, A., Lisowska, B., and Leak, D.J., 2015. The Genus Geobacillus and Their Biotechnological Potential. *Advances in Applied Microbiology*, 92, 1–48.

Hyatt, D.C., Youn, B., Zhao, Y., Santhamma, B., Coates, R.M., Croteau, R.B., and Kang, C., 2007. Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), 5360–5365.

Ignea, C., Cvetkovic, I., Loupassaki, S., Kefalas, P., Johnson, C.B., Kampranis, S.C., and Makris, A.M., 2011. Improving yeast strains using recyclable integration cassettes, for the production of plant terpenoids. *Microbial Cell Factories*, 10, 1–18.

Ikeda, C., Hayashi, Y., Itoh, N., Seto, H., and Dairi, T., 2007. Functional Analysis of Eubacterial ent-Copalyl Diphosphate Synthase and Pimara-9(11),15-Diene Synthase with Unique Primary Sequences. *The Journal of Biochemistry*, 141(1), 37–45.

Ikeda, H., Shin-ya, K., and Omura, S., 2014. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of Industrial Microbiology & Biotechnology*, 41(2), 233–250.

Iyer, P. V and Ananthanarayan, L., 2008. Enzyme stability and stabilization-Aqueous and non-aqueous environment. *Process Biochemistry*, 43, 1019–1032.

Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F., 2001. Protein Flexibility Predictions Using Graph Theory. *Proteins: Structure, Function, and Genetics*, 44(2), 150–165.

Janke, R., Görner, C., Hirte, M., Brück, T., and Loll, B., 2014. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallographica Section D Biological Crystallography*, 70(6), 1528–1537.

Jansen, M.L.A. and van Gulik, W.M., 2014. Towards large scale fermentative production of succinic acid. *Current Opinion in Biotechnology*, 30, 190–197.

Jegannathan, K.R. and Nielsen, P.H., 2013. Environmental assessment of enzyme use in industrial production – a literature review. *Journal of Cleaner Production*, 42, 228–240.

Jha, P., Singh, S., Raghuram, M., Nair, G., Jobby, R., Gupta, A., and Desai, N., 2019. Valorisation of orange peel: supplement in fermentation media for ethanol production and source of limonene. *Environmental Sustainability*, 2(1), 33–41.

Jia, Q., Li, G., Köllner, T.G., Fu, J., Chen, X., Xiong, W., Crandall-Stotler, B.J., Bowman, J.L., Weston, D.J., Zhang, Y., Chen, L., Xie, Y., Li, F.-W., Rothfels, C.J., Larsson, A., Graham, S.W., Stevenson, D.W., Wong, G.K.-S., Gershenzon, J., and Chen, F., 2016. Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *PNAS*, 113(43), 12328–12333.

Jia, Y., Xia, D., and Louzada, E.S., 2005. Molecular Cloning and Expression Analysis of a Putative Terpene Synthase Gene from Citrus. *Journal of the American Society for Horticultural Science*, 130(3), 454–458.

Jiang, J. and Cane, D.E., 2008. Geosmin Biosynthesis. Mechanism of the Fragmentation-Rearrangement in the Conversion of Germacradienol to Geosmin. *Journal of the American Chemical Society*, 130(2), 428–429.

Jiang, J., He, X., and Cane, D.E., 2007. Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. *Nature Chemical Biology*, 3(11), 711–715.

Jiang, Z., Kempinski, C., and Chappell, J., 2016. Extraction and Analysis of Terpenes/Terpenoids. *Current Protocols in Plant Biology*, 1, 345–358.

Jin, X., Xu, L.-H., Mao, P.-H., Hseu, T.-H., and Jiang, C.-L., 1998. Description of Saccharomonospora xinjiangensis sp. nov. based on chemical and molecular classification. *International Journal of Systematic Bacteriology*, 48(4), 1095–1099.

Jochens, H., Aerts, D., and Bornscheuer, U.T., 2010. Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein Engineering, Design and Selection*, 23(12), 903–909.

John, I., Muthukumar, K., and Arunagiri, A., 2017. A review on the potential of citrus

waste for D-Limonene, pectin, and bioethanol production. *International Journal of Green Energy*, 14(7), 599–612.

Johnson, L.B., Gintner, L.P., Park, S., and Snow, C.D., 2015. Discriminating between stabilizing and destabilizing protein design mutations via recombination and simulation. *Protein Engineering, Design and Selection*, 28(8), 259–267.

Johnson, S.R., Bhat, W.W., Sadre, R., Miller, G.P., Garcia, A.S., and Hamberger, B., 2019. Promiscuous terpene synthases from Prunella vulgaris highlight the importance of substrate and compartment switching in terpene synthase evolution. *New Phytologist*, 223(1), 323–335.

Jones, A.M.P., Shukla, M.R., Sherif, S.M., Brown, P.B., and Saxena, P.K., 2016. Growth regulating properties of isoprene and isoprenoid-based essential oils. *Plant Cell Reports*, 35(1), 91–102.

Jones, C.G., Keeling, C.I., Ghisalberti, E.L., Barbour, E.L., Plummer, J.A., and Bohlmann, J., 2008. Isolation of cDNAs and functional characterisation of two multi-product terpene synthase enzymes from sandalwood, Santalum album L. *Archives of Biochemistry and Biophysics*, 477(1), 121–130.

Jones, C.G., Moniodis, J., Zulak, K.G., Scaffidi, A., Plummer, J.A., Ghisalberti, E.L., Barbour, E.L., and Bohlmann, J., 2011. Sandalwood Fragrance Biosynthesis Involves Sesquiterpene Synthases of Both the Terpene Synthase (TPS)-a and TPS-b Subfamilies, including Santalene Synthases. *The Journal of Biological Chemistry*, 286(20), 17445–17454.

Jones, R.V.H. and Sutherland, M.D., 1968. Hedycaryol, the Precursor of Elemol. *Chemical Communications*, (20), 1229–1230.

Jongedijk, E., Cankar, K., Buchhaupt, M., Schrader, J., Bouwmeester, H., and Beekwilder, J., 2016. Biotechnological production of limonene in microorganisms. *Applied Microbiology and Biotechnology*, 100(7), 2927–2938.

Juturu, V. and Wu, J.C., 2016. Microbial production of lactic acid: the latest development. *Critical Reviews in Biotechnology*, 36(6), 967–977.

Kale, V., 2013. Litorilinea aerophila gen. nov., sp. nov., an aerobic member of the class Caldilineae, phylum Chloroflexi, isolated from an intertidal hot spring. *International Journal of Systematic and Evolutionary Microbiology*, 63, 1149–1154.

Kallscheuer, N., Classen, T., Drepper, T., and Marienhagen, J., 2019. Production of plant metabolites with applications in the food industry using engineered microorganisms. *Current Opinion in Biotechnology*, 56, 7–17.

Kamatou, G.P.P., Vermaak, I., Viljoen, A.M., and Lawrence, B.M., 2013. Menthol: A simple monoterpene with remarkable biological properties. *Phytochemistry*, 96, 15–25.

Kang, M.-K., Eom, J.-H., Kim, Y., Um, Y., and Woo, M.H., 2014. Biosynthesis of pinene from glucose using metabolically-engineered Corynebacterium glutamicum. *Biotechnology Letters*, 36(10), 2069–2077.

Kappers, I.F., Aharoni, A., van Herpen, T.W.J.M., Luckerhoff, L.L.P., Dicke, M., and Bouwmeester, H.J., 2005. Genetic Engineering of Terpenoid Metabolism Attracts Bodyguards to Arabidopsis. *Science*, 309(5743), 2070–2072.

Karuppiah, V., Ranaghan, K.E., Leferink, N.G.H., Johannissen, L.O., Shanmugam, M., Cheallaigh, A.N., Bennett, N.J., Kearsey, L.J., Takano, E., Gardiner, J.M., Van Der Kamp, M.W., Hay, S., Mulholland, A.J., Leys, D., and Scrutton, N.S., 2017. Structural Basis of Catalysis in the Bacterial Monoterpene Synthases Linalool Synthase and 1,8-Cineole Synthase. *ACS Catalysis*, 7(9), 6268–6282.

Kawaide, H., Imai, R., Sassa, T., and Kamiya, Y., 1997. ent-Kaurene Synthase from the Fungus Phaeosphaeria sp. L487. *The Journal of Biological Chemistry*, 272(35), 21706–21712.

Kawasaki, T., Hayashi, Y., Kuzuyama, T., Furihata, K., Itoh, N., Seto, H., and Dairi, T., 2006. Biosynthesis of a Natural Polyketide-Isoprenoid Hybrid Compound, Furaquinocin A: Identification and Heterologous Expression of the Gene Cluster. *Journal of Bacteriology*, 188(4), 1236–1244.

Kellogg, E.H., Leaver-Fay, A., and Baker, D., 2010. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins*, 79(3), 830–838.

Keulen, G. Van and Dyson, P.J., 2014. Production of Specialized Metabolites by Streptomyces coelicolorA3(2). *In*: *Advances in Applied Microbiology*. Elsevier Inc., 217–266.

Khanna, C., Rosenberg, M., and Vail, D.M., 2015. A Review of Paclitaxel and Novel Formulations Including Those Suitable for Use in Dogs. *Journal of Veterinary Internal Medicine*, 29(4), 1006–1012.

Kim, B., Sahin, N., Tan, G.Y.A., Zakrzewska-Czerwinska, J., and Goodfellow, M., 2002. Amycolatopsis eurytherma sp. nov., a thermophilic actinomycete isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, 52, 889–894.

Kim, D.-Y., Bochar, D.A., Stauffacher, C. V, and Rodwell, V.W., 1999. Expression and Characterization of the HMG-CoA Reductase of the Thermophilic Archaeon Sulfolobus solfataricus. *Protein Expression and Purification*, 17(3), 435–442.

Kim, E.-M., Woo, H.M., Tian, T., Yilmaz, S., Javidpour, P., Keasling, J.D., and Lee, T.S., 2017. Autonomous control of metabolic state by a quorum sensing (QS)-mediated regulator for bisabolene production in engineered *E. coli*. *Metabolic Engineering*, 44, 325–336.

Kim, M.-S. and Gen Lei, X., 2008. Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR. *Applied Microbiology and Biotechnology*, 79, 69–75.

Kim, M.G., Lee, J.-C., Park, D.-J., Li, W.-J., and Kim, C.-J., 2014. Alicyclobacillus tengchongensis sp. nov., a Thermo-Acidophilic Bacterium Isolated from Hot Spring Soil. *Journal of Microbiology*, 52(10), 884–889.

Kirby, J. and Keasling, J.D., 2008. Metabolic engineering of microorganisms for isoprenoid production. *Natural Product Reports*, 25(4), 656.

Kizer, L., Pitera, D.J., Pfleger, B.F., and Keasling, J.D., 2008. Application of Functional Genomics to Pathway Optimization for Increased Isoprenoid Production. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 74(10), 3229–3241.

Klement, T. and Büchs, J., 2012. Itaconic acid - A biotechnological process in change. *Bioresource Technology*, 135, 422–431.

Köksal, M., Chou, W.K.W., Cane, D.E., and Christianson, D.W., 2012. Structure of 2-Methylisoborneol Synthase from Streptomyces coelicolor and Implications for the Cyclization of a Noncanonical C-Methylated Monoterpenoid Substrate. *Biochemistry*, 51(14), 3011–3020.

Köksal, M., Jin, Y., Coates, R.M., Croteau, R., and Christianson, D.W., 2011. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature*, 469(7328), 116–120.

Köksal, M., Potter, K., Peters, R.J., and Christianson, D.W., 2014. 1.55 Å-Resolution Structure of ent-Copalyl Diphosphate Synthase and Exploration of General Acid Function by Site-Directed Mutagenesis. *Biochim et Biophysica Acta*, 1840(1), 184–190.

Köksal, M., Zimmer, I., Schnitzler, J.-P., and Christianson, D.W., 2010. Structure of Isoprene Synthase Illuminates the Chemical Mechanism of Teragram Atmospheric Carbon Emission. *Journal of Molecular Biology*, 402(2), 363–373.

Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., and Ikeda, H., 2013. Engineered Streptomyces avermitilis Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites. *ACS Synthetic Biology*, 2(7), 384–396.

Komatsu, M., Tsuda, M., Omura, S., Oikawa, H., Ikeda, H., and Ifo, S., 2008. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *PNAS*, 105(21), 7422–7427.

Konstat-Korzenny, E., Ascencio-Aragón, J.A., Niezen-Lugo, S., and Vázquez-López, R., 2018. Artemisinin and Its Synthetic Derivatives as a Possible Therapy for Cancer. *Medical Sciences*, 6(1), 1–10.

Köpke, D., Beyaert, I., Gershenzon, J., Hilker, M., and Schmidt, A., 2010. Species-specific responses of pine sesquiterpene synthases to sawfly oviposition. *Phytochemistry*, 71, 909–917.

Korman, T.P., Opgenorth, P.H., and Bowie, J.U., 2017. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nature Communications*, 8, 1–8.

Koyama, T., Obata, S., Osabe, M., Takeshita, A., Yokoyama, K., Uchida, M., Nishino, T., and Ogura, K., 1993. Thermostable Farnesyl Diphosphate Synthase of Bacillus stearothermophilus: Molecular Cloning, Sequence Determination, Overproduction, and Purification. *The Journal of Biochemistry*, 113(3), 355–363.

Kramer, R. and Abraham, W.-R., 2012. Volatile sesquiterpenes from fungi: what are they good for? *Phytochemistry Reviews*, 11(1), 15–37.

Krug, C., Cordeiro, G.D., Schäffler, I., Silva, C.I., Oliveira, R., Schlindwein, C., Dötterl, S., and Alves-dos-Santos, I., 2018. Nocturnal Bee Pollinators Are Attracted to Guarana Flowers by Their Scents. *Frontiers in Plant Science*, 9, 1–6.

Kumar, P. and Satyanarayana, T., 2009. Critical Reviews in Biotechnology Microbial glucoamylases: characteristics and applications Microbial glucoamylases: characteristics and applications. *Critical Reviews in Biotechnology*, 29(3), 225–255.

Kumar, S., Gupta, S.K., Singh, P., Bajpai, P., Gupta, M.M., Singh, D., Gupta, A.K., Ram, G., Shasany, A.K., and Sharma, S., 2004. High yields of artemisinin by multi-harvest of Artemisia annua crops. *Industrial Crops and Products*, 19(1), 77–90.

Kumar, Y., Khan, F., Rastogi, S., and Shasany, A.K., 2018. Genome-wide detection of terpene synthase genes in holy basil (Ocimum sanctum L.). *PLoS ONE*, 131(11), 1–25.

Kumari, S., Pundhir, S., Priya, P., Jeena, G., Punetha, A., Chawla, K., Jafaree, Z.F., Mondal, S., and Yadav, G., 2014. EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom. *Database*, 1–12.

Lackus, N.D., Lackner, S., Gershenzon, J., Unsicker, S.B., and Köllner, T.G., 2018. The occurrence and formation of monoterpenes in herbivore-damaged poplar roots. *Scientific Reports*, 8, 1–13.

Landmann, C., Fink, B., Festner, M., Dregus, M., Engel, K.-H., and Schwab, W., 2007. Cloning and functional characterization of three terpene synthases from lavender (Lavandula angustifolia). *Archives of Biochemistry and Biophysics*, 465(2), 417–429.

Lanzotti, V., 2013. Diterpenes for Therapeutic Use. *In*: *Natural Products*. 3173–3191.

Laothawornkitkul, J., Paul, N.D., Vickers, C.E., Possell, M., Taylor, J.E., Mullineaux, P.M., and Hewitt, C.N., 2008. Isoprene emissions influence herbivore feeding decisions. *Plant, Cell and Environment*, 31(10), 1410–1415.

Lauchli, R., Rabe, K.S., Kalbarczyk, K.Z., Tata, A., Heel, T., Kitto, R.Z., and Arnold, F.H., 2013. High-Throughput Screening for Terpene-Synthase-Cyclization Activity and Directed Evolution of a Terpene Synthase. *Angewandte Chemie International Edition*, 52(21), 5571–5574.

Leavell, M.D., Mcphee, D.J., and Paddon, C.J., 2016. Developing fermentative terpenoid production for commercial usage. *Current Opinion in Biotechnology*, 37, 114–119.

Lee, K.C.-Y., Dunfield, P.F., Morgan, X.C., Crowe, M.A., Houghton, K.M., Vyssotski, M., Ryan, J.L.J., Lagutin, K., McDonald, I.R., and Stott, M.B., 2011. Chthonomonas caldirosea gen. nov., sp. nov., an aerobic, pigmented, thermophilic micro-organism of a novel bacterial class, Chthonomonadetes classis nov., of the newly described phylum Armatimonadetes originally designated candidate division OP10. *International Journal of Systematic and Evolutionary Microbiology*, 61(10), 2482–2490.

Lee, S. and Poulter, C.D., 2008. Cloning, Solubilization, and Characterization of Squalene Synthase from Thermosynechococcus elongatus BP-1. *Journal of Bacteriology*, 190(11), 3808–3816.

Lehka, B.J., Eichenberger, M., Emil Bjørn-Yoshimoto, W., Vanegas, K.G., Buijs, N., Jensen, N.B., Dyekjaer, J.D., Jenssen, H., Simon, E., and Naesby, M., 2017. Improving heterologous production of phenylpropanoids in Saccharomyces cerevisiae by tackling an unwanted side reaction of Tsc13, an endogenous double-bond reductase. *FEMS Yeast Research*, 17(1), 1–12.

Lehmann, M., Pasamontes, L., Lassen, S.F., and Wyss, M., 2000. The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1543(2), 408–415.

Lehmann, M. and Wyss, M., 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology*, 12(4), 371–375.

Lesburg, C.A., Zhai, G., Cane, D.E., and Christianson, D.W., 1997. Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology. *Science*, 277(5333), 1820–1824.

Li, J., Qin, S., You, Z.-Q., Long, L.-J., Tian, X.-P., Wang, F.-Z., and Zhang, S., 2013. Melghirimyces profundicolus sp. nov., isolated from a deep-sea sediment. *International Journal of Systematic and Evolutionary Microbiology*, 63, 4552–4556.

Li, M., Nian, R., Xian, M., and Zhang, H., 2018. Metabolic engineering for the production of isoprene and isopentenol by Escherichia coli. *Applied Microbiology and Biotechnology*, 102(18), 7725–7738.

Li, Q., Fan, F., Gao, X., Yang, C., Bi, C., Tang, J., Liu, T., and Zhang, X., 2017. Balanced activation of IspG and IspH to eliminate MEP intermediate

accumulation and improve isoprenoids production in *Escherichia coli*. *Metabolic Engineering*, 44, 13–21.

Li, R., Chou, W.K.W., Himmelberger, J.A., Litwin, K.M., Harris, G.G., Cane, D.E., and Christianson, D.W., 2014. Reprogramming the Chemodiversity of Terpenoid Cyclization by Remolding the Active Site Contour of epi-Isozizaene Synthase. *Biochemistry*, 53(7), 1155–1168.

Li, Y., Zhang, G., and Pfeifer, B.A., 2015. Current and Emerging Options for Taxol Production. *Advances in Biochemical Engineering/Biotechnology*, 148, 405–425.

Liang, P.-H., Ko, T.-P., and Wang, A.H.-J., 2002. Structure, mechanism and function of prenyltransferases. *European Journal of Biochemistry*, 269(14), 3339–3354.

Liang, W., Wang, H., Xu, N., Zhou, W., Ju, J., Liu, J., and Ma, Y., 2019. Metabolic engineering of Corynebacterium glutamicum for L-cysteine production. *Applied Microbiology and Biotechnology*, 103(3), 1325–1338.

Lin, P.P., Rabe, K.S., Takasumi, J.L., Kadisch, M., Arnold, F.H., and Liao, J.C., 2014. Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius. *Metabolic Engineering*, 24, 1–8.

Lin, X. and Cane, D.E., 2009. Biosynthesis of the Sesquiterpene Antibiotic Albaflavenone in Streptomyces coelicolor. Mechanism and Stereochemistry of the Enzymatic Formation of Epi-isozizaene. *Journal of the American Chemical Society*, 131(18), 6332–6333.

Lin, X., Hopson, R., and Cane, D.E., 2006. Genome Mining in Streptomyces coelicolor: Molecular Cloning and Characterization of a New Sesquiterpene Synthase. *Journal of the American Chemical Society*, 128(18), 6022–6023.

Liolios, K., Sikorski, J., Jando, M., Lapidus, A., Copeland, A., Glavina, T., Nolan, M., Lucas, S., Tice, H., Cheng, J.-F., Han, C., Woyke, T., Goodwin, L., Pitluck, S., Ivanova, N., Mavromatis, K., Mikhailova, N., Chertkov, O., Kuske, C., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Detter, J.C., Brettin, T., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., and Kyrpides, N.C., 2010. Complete genome sequence of Thermobispora bispora type strain (R51T). *Standards in Genomic Sciences*, 2(3), 318–326.

Liu, C.-L., Tian, T., Alonso-Gutierrez, J., Garabedian, B., Wang, S., Baidoo, E.E.K.,

Benites, V., Chen, Y., Petzold, C.J., Adams, P.D., Keasling, J.D., Tan, T., and Lee, T.S., 2018. Renewable production of high density jet fuel precursor sesquiterpenes from Escherichia coli. *Biotechnology for Biofuels*, 11, 1–15.

Liu, R., Deng, Z., and Liu, T., 2018. Streptomyces species: Ideal chassis for natural product discovery and overproduction. *Metabolic Engineering*, 50, 74–84.

Liu, X., Ding, W., and Jiang, H., 2017. Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. *Microbial Cell Factories*, 16, 1–9.

Liu, Y., Jing, S.-X., Luo, S.-H., and Li, S.-H., 2019. Non-volatile natural products in plant glandular trichomes: chemistry, biological activities and biosynthesis. *Natural Product Reports*, 36(4), 626–665.

Liu, Y. and Kuhlman, B., 2006. RosettaDesign server for protein design. *Nucleic Acids Research*, 34, W235–W238.

Llopiz, P., Jürgens, U.J., and Rohmer, M., 1996. Prokaryotic triterpenoids: Bacteriohopanetetrol glycuronosides from the thermophilic cyanobacterium *Synechococcus* PCC 6907. *FEMS Microbiology Letters*, 140(2–3), 199–202.

Lodewyk, M.W., Gutta, P., and Tantillo, D.J., 2008. Computational Studies on Biosynthetic Carbocation Rearrangements Leading to Sativene, Cyclosativene, α-Ylangene, and β-Ylangene. *The Journal of Organic Chemistry*, 73(17), 6570–6579.

Lombard, J. and Moreira, D., 2010. Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life. *Molecular Biology and Evolution*, 28(1), 87–99.

Lv, X., Wang, F., Zhou, P., Ye, L., Xie, W., Xu, H., and Yu, H., 2016. Dual regulation of cytoplasmic and mitochondrial acetyl-CoA utilization for improved isoprene production in Saccharomyces cerevisiae. *Nature Communications*, 21(7).

Mabrouk, S. Ben, Ayadi, D.Z., Hlima, B., and Bejar, S., 2013. Thermostability improvement of maltogenic amylase MAUS149 by error prone PCR. *Journal of Biotechnology*, 168, 601–606.

Maheshwari, R., Bharadwaj, G., and Bhat, M.K., 2000. Thermophilic Fungi: Their Physiology and Enzymes. *Microbiology and Molecular Biology Reviews*, 64(3), 461–488.

Mamonova, T.B., Glyakina, A. V, Galzitskaya, O. V, and Kurnikova, M.G., 2013. Stability and rigidity/flexibility—Two sides of the same coin? *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(5), 854–866.

Manachini, P.L., Mora, D., Nicastro, G., Parini, C., Stackebrandt, E., Pukall, R., and Fortina, M.G., 2000. Bacillus thermodenitrificans sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology*, 50(3), 1331–1337.

Mandai, T., Kawada, M., and Otera, J., 1983. Highly Stereoselective Total Syntheses of β-Farnesene, β-Sinensal, and Dendrolasin Employing 2-(Hydroxymethyl)-4-(phenylthio)-l-butene as a Building Block. *The Journal of Organic Chemistry*, 48, 5183–5185.

Mann, F.M., Xu, M., Chen, X., Fulton, D.B., Russell, D.G., and Peters, R.J., 2009. Edaxadiene: A New Bioactive Diterpene from Mycobacterium tuberculosis. *Journal of the American Chemical Society*, 131(48), 17526–17527.

Martin, D.M. and Bohlmann, J., 2004. Identification of Vitis vinifera (-)-α-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry*, 65(9), 1223–1229.

Martin, D.M., Fäldt, J., and Bohlmann, J., 2004. Functional Characterization of Nine Norway Spruce TPS Genes and Evolution of Gymnosperm Terpene Synthases of the TPS-d Subfamily. *Plant Physiology*, 135(4), 1908–1927.

Martinez, F.A.C., Balciunas, E.M., Salgado, J.M., González, J.M.D., Converti, A., and de Souza Oliveira, R.P., 2013. Lactic acid properties, applications and production: A review. *Trends in Food Science & Technology*, 30(1), 70–83.

Matsuba, Y., Nguyen, T.T.H., Wiegert, K., Falara, V., Gonzales-Vigil, E., Leong, B., Schäfer, P., Kudrna, D., Wing, R.A., Bolger, A.M., Usadel, B., Tissier, A., Fernie, A.R., Barry, C.S., and Pichersky, E., 2013. Evolution of a Complex Locus for Terpene Biosynthesis in Solanum. *The Plant Cell*, 25(6), 2022–2036.

Mavromatis, K., Ivanova, N., Anderson, I., Lykidis, A., Hooper, S.D., Sun, H., Kunin, V., Lapidus, A., Hugenholtz, P., Patel, B., and Kyrpides, N.C., 2009. Genome Analysis of the Anaerobic Thermohalophilic Bacterium *Halothermothrix orenii*. *PLoS ONE*, 4(1), 4192.

Mavromatis, K., Sikorski, J., Lapidus, A., Glavina Del Rio, T., Copeland, A., Tice, H., Cheng, J.-F., Lucas, S., Chen, F., Nolan, M., Bruce, D., Goodwin, L., Pitluck, S.,

Ivanova, N., Ovchinnikova, G., Pati, A., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Chain, P., Meincke, L., Sims, D., Chertkov, O., Han, C., Brettin, T., Detter, J.C., Wahrenburg, C., Rohde, M., Pukall, R., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., and Kyrpides, N.C., 2010. Complete genome sequence of *Alicyclobacillus acidocaldarius* type strain (104-IA^T). *Standards in Genomic Sciences*, 2(1), 9–18.

McCarthy, A.J. and Cross, T., 1984. A Taxonomic Study of Thermomonospora and Other Monosporic Actinomycetes. *Journal of' General Microbiologv*, 130, 5–25.

McCormick, A.C., Unsicker, S.B., and Gershenzon, J., 2012. The specificity of herbivore-induced plant volatiles in attracting herbivore enemies. *Trends in Plant Science*, 17(5), 303–310.

McCullum, E.O., Williams, B.A.R., Zhang, J., and Chaput, J.C., 2010. Random Mutagenesis by Error-Prone PCR. *In*: *In Vitro Mutagenesis Protocols*. 103–110.

McPhee, D., Pin, A., Kizer, L., and Perelman, L., 2014. Deriving Renewable Squalene from Sugarcane. *Cosmetic Toiletries*, 129(6).

Meadows, A.L., Hawkins, K.M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R.H., Tai, A., Mahatdejkul-Meadows, T., Xu, L., Zhao, L., Dasika, M.S., Murarka, A., Lenihan, J., Eng, D., Leng, J.S., Liu, C.-L., Wenger, J.W., Jiang, H., Chao, L., Westfall, P., Lai, J., Ganesan, S., Jackson, P., Mans, R., Platt, D., Reeves, C.D., Saija, P.R., Wichmann, G., Holmes, V.F., Benjamin, K., Hill, P.W., Gardner, T.S., and Tsong, A.E., 2016. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, 537, 694–697.

Meeks, J.C. and Castenholz, R.W., 1971. Growth and Photosynthesis in an Extreme Thermophile, Synechococcus Lividus (Cyanophyta). *Archives of Microbiology*, 78, 25–41.

van Der Meer, M.T.J., Klatt, C.G., Wood, J., Bryant, D.A., Bateson, M.M., Lammerts, L., Schouten, S., Damsté, J.S.S., Madigan, M.T., and Ward, D.M., 2010. Cultivation and Genomic, Nutritional, and Lipid Biomarker Characterization of Roseiflexus Strains Closely Related to Predominant In Situ Populations Inhabiting Yellowstone Hot Spring Microbial Mats. *Journal of Bacteriology*, 192(12), 3033–3042.

Meguro, A., Tomita, T., Nishiyama, M., and Kuzuyama, T., 2013. Identification and

Characterization of Bacterial Diterpene Cyclases that Synthesize the Cembrane Skeleton. *ChemBioChem*, 14(3), 316–321.

Mehta, B.J., Obraztsova, I.N., and Cerdá-Olmedo, E., 2003. Mutants and Intersexual Heterokaryons of Blakeslea trispora for Production of-Carotene and Lycopene. *Applied and Environmental Microbiology*, 69(7), 4043–4048.

Mei, Y.-Z., Liu, R.-X., Wang, D.-P., Wang, X., and Dai, C.-C., 2015. Biocatalysis and biotransformation of resveratrol in microorganisms. *Biotechnology Letters*, 37(1), 9–18.

Memari, H.R., Pazouki, L., and Niinemets, Ü., 2013. The Biochemistry and Molecular Biology of Volatile Messengers in Trees. *In*: *Biology, Controls and Models of Tree Volatile Organic Compound Emissions*. 47–94.

Mewalal, R., Rai, D.K., Kainer, D., Chen, F., Külheim, C., Peter, G.F., and Tuskan, G.A., 2017. Plant-Derived Terpenes: A Feedstock for Specialty Biofuels. *Trends in Biotechnology*, 35(3), 227–240.

Meylemans, H.A., Quintana, R.L., and Harvey, B.G., 2012. Efficient conversion of pure and mixed terpene feedstocks to high density fuels. *Fuel*, 97, 560–568.

Mierziak, J., Kostyn, K., and Kulma, A., 2014. Flavonoids as Important Molecules of Plant Interactions with the Environment. *Molecules*, 19(10), 16240–16265.

Miller, S.R., Castenholz, R.W., and Pedersen, D., 2007. Phylogeography of the Thermophilic Cyanobacterium Mastigocladus laminosus. *Applied and Environmental Microbiology*, 73(15), 4751–4759.

Minerdi, D., Bossi, S., Maffei, M.E., Gullino, M.L., and Garibaldi, A., 2011. Fusarium oxysporum and its bacterial consortium promote lettuce growth and expansin A5 gene expression through microbial volatile organic compound (MVOC) emission. *FEMS Microbiology Ecology*, 76(2), 342–351.

Miziorko, H.M., 2011. Enzyme of the mevalonate pathway of isoprenoid biosynthesis. *Archives of Biochemistry and Biophysics*, 505(2), 131–143.

Modarres, H.P., Mofrad, M.R., and Sanati-Nezhad, A., 2016. Protein thermostability engineering. *RSC Advances*, 6(116), 115252–115270.

Mohagheghi, A., Grohmann, K., Himmel, M., Leighton, L., and Updegraff, D.M., 1986. Isolation and Characterization of Acidothermus cellulolyticus gen. nov., sp. nov., a New Genus of Thermophilic, Acidophilic, Cellulolytic Bacteria. *International*

Journal of Systematic Bacteriology, 36(3), 435–443.

Mohammadi, S., Pol, A., van Alen, T.A., Jetten, M.S., and Op den Camp, H.J., 2016. Methylacidiphilum fumariolicum SolV, a thermoacidophilic 'Knallgas' methanotroph with both an oxygen-sensitive and -insensitive hydrogenase. *The ISME Journal*, 11(4), 945–958.

Morehouse, B.R., Kumar, R.P., Matos, J.O., Olsen, S.N., Entova, S., and Oprian, D.D., 2017. Functional and Structural Characterization of a (+)-Limonene Synthase from Citrus sinensis. *Biochemistry*, 56(12), 1706–1715.

Morrone, D., Chambers, J., Lowry, L., Kim, G., Anterola, A., Bender, K., and Peters, R.J., 2009. Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in Bradyrhizobium japonicum. *FEBS Letters*, 583(2), 475–480.

Mougiakos, I., Mohanraju, P., Bosma, E.F., Vrouwe, V., Bou, M.F., Naduthodi, M.I.S., Gussak, A., Brinkman, R.B.L., van Kranenburg, R., and van Der Oost, J., 2017. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature Communications*, 8, 1–11.

Munné-Bosch, S., Mueller, M., Schwarz, K., and Alegre, L., 2001. Diterpenes and antioxidative protection in drought-stressed Salvia officinalis plants. *Journal of Plant Physiology*, 158(11), 1431–1437.

Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G., and Dudareva, N., 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *The Plant Journal*, 55(2), 224–239.

Nagel, R., Turrini, P.C.G., Nett, R.S., Leach, J.E., Verdier, V., van Sluys, M.A., and Peters, R.J., 2017. An operon for production of bioactive gibberellin A4 phytohormone with wide distribution in the bacterial rice leaf streak pathogen Xanthomonas oryzae pv. oryzicola. *New Phytologist*, 214(3), 1260–1266.

Nakagawa, A., Minami, H., Kim, J.-S., Koyanagi, T., Katayama, T., Sato, F., and Kumagai, H., 2011. A bacterial platform for fermentative production of plant alkaloids. *Nature Communications*, 2, 1–8.

Nakamura, L.K., Blumenstock, I., and Claus, D., 1988. Taxonomic Study of Bacillus coagulans Hammer 1915 with a Proposal for Bacillus smithii sp. nov. *International Journal of Systematic Bacteriology*, 38(1), 63–73.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S., 2002. Complete Genome Structure of the Thermophilic Cyanobacterium Thermosynechococcus elongatus BP-1. *DNA Research*, 9, 123–130.

Nakano, C. and Hoshino, T., 2009. Characterization of the Rv3377c Gene Product , a Type-B Diterpene Cyclase , from the Mycobacterium tuberculosis H37 Genome. *ChemBioChem*, 10(12), 2060–2071.

Nakano, C., Kim, H.-K., and Ohnishi, Y., 2011a. Identification and Characterization of the Linalool/Nerolidol Synthase from Streptomyces clavuligerus. *ChemBioChem*, 12(16), 2403–2407.

Nakano, C., Kim, H.-K., and Ohnishi, Y., 2011b. Identification of the First Bacterial Monoterpene Cyclase, a 1,8-Cineole Synthase, that Catalyzes the Direct Conversion of Geranyl Diphosphate. *ChemBioChem*, 12(13), 1988–1991.

Nakano, C., Kudo, F., Eguchi, T., and Ohnishi, Y., 2011. Genome Mining Reveals Two Novel Bacterial Sesquiterpene Cyclases: (−)-Germacradien-4-ol and (−)-epi-α-Bisabolol Synthases from Streptomyces citricolor. *ChemBioChem*, 12(15), 2271–2275.

Nakano, C., Ootsuka, T., Takayama, K., Mitsui, T., Sato, T., and Hoshino, T., 2011. Characterization of the Rv3378c Gene Product, a New Diterpene Synthase for Producing Tuberculosinol and (13R, S)-Isotuberculosinol (Nosyberkol), from the Mycobacterium tuberculosis H37Rv Genome. *Bioscience, Biotechnology, and Biochemistry*, 75(1), 75–81.

Nakano, C., Oshima, M., Kurashima, N., and Hoshino, T., 2015. Identification of a New Diterpene Biosynthetic Gene Cluster that Produces O-Methylkolavelool in Herpetosiphon aurantiacus. *ChemBioChem*, 16(5), 772–781.

Nakano, C., Tezuka, T., Horinouchi, S., and Ohnishi, Y., 2012. Identification of the SGR6065 gene product as a sesquiterpene cyclase involved in (+)-epicubenol biosynthesis in Streptomyces griseus. *The Journal of Antibiotics*, 65(11), 551–558.

Nalin, R., Putra, S.R., Domenach, A.-M., Rohmer, M., Gourbiere, F., and Berry, A.M., 2000. High hopanoid/total lipids ratio in Frankia mycelia is not related to the

nitrogen status. *Microbiology*, 146(11), 3013–3019.

Nariman Addou, A., Schumann, P., Spröer, C., Bouanane-Darenfed, A., Amarouche-Yala, S., Hacene, H., Cayol, J.-L., and Fardeau, M.-L., 2013. Melghirimyces thermohalophilus sp. nov., a thermoactinomycete isolated from an Algerian salt lake. *International Journal of Systematic and Evolutionary Microbiology*, 63(5), 1717–1722.

Nartey, C.M., Aljadeff, J., Fernandez, I., Laurendon, C., Koo, H.J., O'Maille, P.E., Sharpee, T.O., and Noel, J.P., 2017. Quantifying the thermostability landscape separating plant sesquiterpene synthase orthologs using Maximum Entropy. *bioRxiv*, 1–25.

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E. V, Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V. V, Osipov, G.A., Belyaev, S.S., and Ivanov, M. V, 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of Geobacillus subterraneus gen. nov., sp. nov. and Geobacillus uzenensis sp. nov. from petroleum reservoirs. *International Journal of Systematic and Evolutionary Microbiology*, 51(2), 433–446.

Ndaba, B., Chiyanzu, I., and Marx, S., 2015. n-Butanol derived from biochemical and chemical routes: A review. *Biotechnology Reports*, 8, 1–9.

Nerio, L.S., Olivero-Verbel, J., and Stashenko, E., 2010. Repellent activity of essential oils: A review. *Bioresource Technology*, 101(1), 372–378.

Nett, R.S., Montanares, M., Marcassa, A., Lu, X., Nagel, R., Charles, T.C., Hedden, P., Rojas, M.C., and Peters, R.J., 2017. Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nature Chemical Biology*, 13(1), 69–74.

Newman, D.J. and Cragg, G.M., 2012. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products*, 75(3), 311–335.

Nicolaou, K.C., Yang, Z., Liu, J.J., Ueno, H., Nantermet, P.G., Guy, R.K., Clalborne, C.F., Renaud, J., Couladouros, E.A., Paulvannan, K., and J, S.E., 1994. Total synthesis of taxol. *Nature*, 367, 630–634.

Nieuwenhuizen, N.J., Wang, M.Y., Matich, A.J., Green, S.A., Chen, X., Yauk, Y.-K., Beuning, L.L., Nagegowda, D.A., Dudareva, N., and Atkinson, R.G., 2009. Two terpene synthases are responsible for the major sesquiterpenes emitted from

the flowers of kiwifruit (Actinidia deliciosa). *Journal of Experimental Botany*, 60(11), 3203–3219.

Nishimura, H., Azami, Y., Miyagawa, M., Hashimoto, C., Yoshimura, T., and Hemmi, H., 2013. Biochemical evidence supporting the presence of the classical mevalonate pathway in the thermoacidophilic archaeon Sulfolobus solfataricus. *The Journal of Biochemistry*, 153(5), 415–420.

Niu, C., Zhu, L., Xu, X., and Li, Q., 2016. Rational Design of Disulfide Bonds Increases Thermostability of a Mesophilic 1,3-1,4-β-Glucanase from Bacillus terquilensis. *PLoS ONE*, 11(4), 1–17.

O'Maille, P.E., Malone, A., Dellas, N., Hess Jr., B.A., Smentek, L., Sheehan, I., Greenhagen, B.T., Chappell, J., Manning, G., and Noel, J.P., 2008. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature Chemical Biology*, 4(10), 617–623.

Oikawa, H., Toyomasu, T., Toshima, H., Ohashi, S., Kawaide, H., Kamiya, Y., Ohtsuka, M., Shinoda, S., Mitsuhashi, W., and Sassa, T., 2001. Cloning and Functional Expression of cDNA Encoding Aphidicolan-16β-ol Synthase: A Key Enzyme Responsible for Formation of an Unusual Diterpene Skeleton in Biosynthesis of Aphidicolin. *Journal of American Chemical Society*, 123(21), 5154–5155.

Olsson, K., Keis, S., Morgan, H.W., Dimroth, P., and Cook, G.M., 2003. Bioenergetic Properties of the Thermoalkaliphilic Bacillus sp. Strain TA2.A1. *Journal of Bacteriology*, 185(2), 461–465.

Orsi, E., Folch, P.L., Monje-López, V.T., Fernhout, B.M., Turcato, A., Kengen, S.W.M., Eggink, G., and Weusthuis, R.A., 2019. Characterization of heterotrophic growth and sesquiterpene production by Rhodobacter sphaeroides on a defined medium. *Journal of Industrial Microbiology & Biotechnology*, 46(8), 1179–1190.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Igbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling,

J.D., Reiling, K.K., Renninger, N.S., and Newman, J.D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446), 528–532.

Palmisano, M.M., Nakamura, L.K., Duncan, K.E., Istock, C.A., and Cohan, F.M., 2001. Bacillus sonorensis sp. nov., a close relative of Bacillus licheniformis, isolated from soil in the Sonoran Desert, Arizona. *International Journal of Systematic and Evolutionary Microbiology*, 51, 1671–1679.

Paramasivan, K. and Mutturi, S., 2017. Progress in terpene synthesis strategies through engineering of Saccharomyces cerevisiae. *Critical Reviews in Biotechnology*, 37(8), 974–989.

Pati, A., Labutti, K., Pukall, R., Nolan, M., Glavina, T., Rio, D., Tice, H., Cheng, J.-F., Lucas, S., Chen, F., Copeland, A., Ivanova, N., Mavromatis, K.-T., Mikhailova, N., Pitluck, S., Bruce, D., Goodwin, L., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Chen, A., Palaniappan, K.-N., Chain, P., Brettin, T., Sikorski, J., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.-P., and Lapidus, A., 2010. Complete genome sequence of Sphaerobacter thermophilus type strain (S 6022T). *Standards in Genomic Sciences*, 2(1), 49–56.

Pati, A., Sikorski, J., Nolan, M., Lapidus, A., Copeland, A., Del Rio, T.G., Lucas, S., Chen, F., Tice, H., Pitluck, S., Cheng, J.-F., Chertkov, O., Brettin, T., Han, C., Detter, J.C., Kuske, C., Bruce, D., Goodwin, L., Chain, P., D'haeseleer, P., Chen, A., Palaniappan, K., Ivanova, N., Mavromatis, K., Mikhailova, N., Rohde, M., Tindall, B.J., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., and Klenk, H.-P., 2009. Complete genome sequence of Saccharomonospora viridis type strain (P101T). *Standards in Genomic Sciences*, 1(2), 141–149.

Pazouki, L. and Niinemetst, Ü., 2016. Multi-Substrate Terpene Synthases: Their Occurrence and Physiological Significance. *Frontiers in Plant Science*, 7, 1–16.

Pelot, K.A., Chen, R., Hagelthorn, D.M., Young, C.A., Addison, J.B., Muchlinski, A., Tholl, D., and Zerbe, P., 2018. Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. *Plant Physiology*, 178(1), 54–71.

Pemberton, T.A., Chen, M., Harris, G.G., Chou, W.K.W., Duan, L., Köksal, M., Genshaft, A.S., Cane, D.E., and Christianson, D.W., 2017. Exploring the Influence of Domain Architecture on the Catalytic Function of Diterpene

Synthases. *Biochemistry*, 56(14), 2010–2023.

Peralta-yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., and Lee, T.S., 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications*, 2(483), 1–8.

Pérez-Gill, J. and Rodríguez-Concepción, M., 2013. Metabolic plasticity for isoprenoid biosynthesis in bacteria. *Biochemical Journal*, 452(1), 19–25.

Peters, R.J., 2010. Two rings in them all: The labdane-related diterpenoids. *Natural Product Reports*, 27(11), 1521–1530.

Peters, R.J., Ravn, M.M., Coates, R.M., and Croteau, R.B., 2001. Bifunctional Abietadiene Synthase: Free Diffusive Transfer of the (+)-Copalyl Diphosphate Intermediate between Two Distinct Active Sites. *Journal of American Chemical Society*, 123(37), 8974–8978.

Picaud, S., Brodelius, M., and Brodelius, P.E., 2005. Expression, purification and characterization of recombinant (E)-β-farnesene synthase from Artemisia annua. *Phytochemistry*, 66(9), 961–967.

Pinheiro Alves de Souza, Y., Faria da Mota, F., and Soares Rosado, A., 2017. Draft Genome Sequence of Geobacillus sp. LEMMY01, a Thermophilic Bacterium Isolated from the Site of a Burning Grass Pile. *Genome Announcements*, 5(19).

Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metabolic Engineering*, 9(2), 193–207.

Poli, A., Guven, K., Romano, I., Pirinccioglu, H., Guven, R.G., Euzeby, J.P.M., Matpan, F., Acer, O., Orlando, P., and Nicolaus, B., 2013. Geobacillus subterraneus subsp. aromaticivorans subsp. nov., a novel thermophilic and alkaliphilic bacterium isolated from a hot spring in Sırnak, Turkey. *The Journal of General and Applied Microbiology*, 58(6), 437–446.

Poppe, L., Novbk, L., and Szántaya, C., 1987. A convenient synthesis of (E)-β-farnesene. *Synthetic Communications*, 17(2), 173–179.

Poralla, K., Härtner, T., and Kannenberg, E., 1984. Effect of temperature and pH on the hopanoid content of Bacillus acidocaldarius. *FEMS Microbiology Letters*, 23(2–3), 253–256.

Potter, K., Criswell, J., Zi, J., Stubbs, A., and Peters, R.J., 2014. Novel Product

Chemistry from Mechanistic Analysis of ent-Copalyl Diphosphate Synthases from Plant Hormone Biosynthesis. *Angewandte Chemie International Edition*, 53(28), 7198–7202.

Potter, K.C., Zi, J., Hong, Y.J., Schulte, S., Malchow, B., Tantillo, D.J., and Peters, R.J., 2016. Blocking Deprotonation with Retention of Aromaticity in aP lant ent-Copalyl Diphosphate Synthase Leads to Product Rearrangement. *Angewandte Chemie International Edition*, 55(2), 634–638.

Prisic, S., Xu, J., Coates, R.M., and Peters, R.J., 2007. Probing the Role of the DXDD Motif in Class II Diterpene Cyclases. *ChemBioChem*, 8(8), 869–874.

Priya, P., Kumari, S., and Yadav, G., 2016. Quantification of the plant terpenome: predicted versus actual emission potentials. *Indian Society for Plant Physiology*, 21(4), 569–575.

Pucci, F., Bourgeas, R., and Rooman, M., 2016. Predicting protein thermal stability changes upon point mutations using statistical potentials: Introducing HoTMuSiC. *Nature Publishing Group*.

Pulido, P., Perello, C., and Rodriguez-Concepcion, M., 2012. New Insights into Plant Isoprenoid Metabolism. *Molecular Plant*, 5(5), 964–967.

Qin, Z., Devine, R., Hutchings, M.I., and Wilkinson, B., 2019. A role for antibiotic biosynthesis monooxygenase domain proteins in fidelity control during aromatic polyketide biosynthesis. *Nature Communications*, 10(1), 1–10.

Rabe, P. and Dickschat, J.S., 2013. Rapid Chemical Characterization of Bacterial Terpene Synthases. *Angewandte Chemie International Edition*, 52(6), 1810–1812.

Rabe, P., Pahirulzaman, K.A.K., and Dickschat, J.S., 2015. Structures and Biosynthesis of Corvol Ethers - Sesquiterpenes from the Actinomycete Kitasatospora setae. *Angewandte Chemie International Edition*, 54(20), 6041–6045.

Rabe, P., Schmitz, T., and Dickschat, J.S., 2016. Mechanistic investigations on six bacterial terpene cyclases. *Beilstein Journal of Organic Chemistry*, 12, 1839–1850.

Rahman, R.N.Z.R.A., Leow, T.C., Salleh, A.B., and Basri, M., 2007. Geobacillus zalihae sp. nov., a thermophilic lipolytic bacterium isolated from palm oil mill

effluent in Malaysia. *BMC Microbiology*, 7, 1–10.

Rasmann, S., Köllner, T.G., Degenhardt, J., Hiltpold, I., Toepfer, S., Kuhlmann, U., Gershenzon, J., and Turlings, T.C.J., 2005. Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*, 434(7034), 732–737.

Reetz, M.T. and Carballeira, J.D., 2006. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angewandte Chemie International Edition*, 45, 7745–7751.

Reetz, M.T. and Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols*, 2(4), 891–903.

Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D.J., and Ellis, T., 2016. The Geobacillus Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology*, 5(12), 1342–1347.

Reijula, K.E., 1993. Two bacteria causing farmer's lung: Fine structure of Thermoactinomyces vulgaris and Saccharopolyspora rectivirgula. *Mycopathologia*, 121(3), 143–147.

Ridley, R.G., 2002. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature*, 415, 686–693.

Rigoldi, F., Donini, S., Redaelli, A., Parisini, E., and Gautieri, A., 2018. Engineering of thermostable enzymes for industrial applications. *APL Bioengineering*, 2(1), 1–17.

Rosano, G.L. and Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, 5, 1–17.

Rudolf, J.D., Dong, L.-B., Cao, H., Hatzos-Skintges, C., Osipiuk, J., Endres, M., Chang, C.-Y., Ma, M., Babnigg, G., Joachimiak, A., Phillips, G.N., and Shen, B., 2016. Structure of the ent-Copalyl Diphosphate Synthase PtmT2 from Streptomyces platensis CB00739, a Bacterial Type II Diterpene Synthase. *Journal of American Chemical Society*, 138(34), 10905–10915.

Ruzicka, L., 1953. The Isoprene Rule and the Biogenesis of Terpenic Compounds. *Experientia*, 9(10), 357–367.

Ruzicka, L., 1959. Faraday Lecture - History of the Isoprene Rule. *Proceedings of the Chemical Society of London*, 341–360.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis,

K.B., and Erlich, H.A., 1988. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, 239(4839), 487–491.

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-y-Terrón, R., and Martínez-Contreras, R.D., 2018. Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. *Microbiological Research*, 208, 85–98.

Salehi, B., Stojanović-Radić, Z., Matejić, J., Sharopov, F., Antolak, H., Kręgiel, D., Sen, S., Sharifi-Rad, M., Acharya, K., Sharifi-Rad, R., Martorell, M., Sureda, A., Martins, N., and Sharifi-Rad, J., 2018. Plants of Genus Mentha: From Farm to Food Factory. *Plants*, 7(3), 1–36.

Salmon, M., Laurendon, C., Vardakou, M., Cheema, J., Defernez, M., Green, S., Faraldos, J.A., and Maille, P.E.O., 2015. Emergence of terpene cyclization in Artemisia annua. *Nature Communications*, 6(6143), 1–10.

Sanchez, S., Rodríguez-Sanoja, R., Ramos, A., and Demain, A.L., 2018. Our microbes not only produce antibiotics, they also overproduce amino acids. *Journal of Antibiotics*, 71(1), 26–36.

Sarmiento, F., Peralta, R., and Blamey, J.M., 2015. Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Frontiers in Bioengineering and Biotechnology*, 3, 1–15.

Sato, T., Yoshida, S., Hoshino, H., Tanno, M., Nakajima, M., and Hoshino, T., 2011. Sesquarterpenes (C35 Terpenes) Biosynthesized via the Cyclization of a Linear C35 Isoprenoid by a Tetraprenyl-β-curcumene Synthase and a Tetraprenyl-β-curcumene Cyclase: Identification of a New Terpene Cyclase. *Journal of the American Chemical Society*, 133(25), 9734–9737.

Sauer, M., Porro, D., Mattanovich, D., and Branduardi, P., 2008. Microbial production of organic acids: expanding the markets. *Trends in Biotechnology*, 26(2), 10–108.

Schäfer, H. and Wink, M., 2009. Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis. *Biotechnology Journal*, 4(12), 1684–1703.

Schalk, M., 2013. Method for producing sclareol. *Patent US 8617860 B2*.

Schalk, M., Pastore, L., Mirata, M.A., Khim, S., Schouwey, M., Deguerry, F., Pineda,

V., Rocci, L., and Daviet, L., 2012. Toward a Biosynthetic Route to Sclareol and Amber Odorants. *Journal of the American Chemical Society*, 134(46), 18900–18903.

Schendel, F.J., Bremmon, C.E., Flickinger, M.C., Guettler, M., and Hanson, R.S., 1990. L-Lysine Production at 50°C by Mutants of a Newly Isolated and Characterized Methylotrophic Bacillus sp. *Applied and Environmental Microbiology*, 56(4), 963–970.

Schmidt, I., Schewe, H., Gassel, S., Jin, C., Buckingham, J., Hümbelin, M., Sandmann, G., and Schrader, J., 2011. Biotechnological production of astaxanthin with Phaffia rhodozyma/Xanthophyllomyces dendrorhous. *Applied Microbiology and Biotechnology*, 89(3), 555–571.

Schnee, C., Köllner, T.G., Gershenzon, J., and Degenhardt, J., 2002. The Maize Gene terpene synthase 1 Encodes a Sesquiterpene Synthase Catalyzing the Formation of (E)-β-Farnesene, (E)-Nerolidol, and (E,E)-Farnesol after Herbivore Damage. *Plant Physiology*, 130(4), 2049–2060.

Schnee, C., Köllner, T.G., Held, M., Turlings, T.C.J., Gershenzon, J., and Degenhardt, J., 2006. The Products of a Single Maize Sesquiterpene Synthase Form a Volatile Defense Signal That Attracts Natural Enemies of Maize Herbivores. *PNAS*, 103(4), 1129–1134.

Schofer, S.J., McPhee, D.J., Moriguchi, N., Yamana, Y., Chapman, B., Hirata, K., and Uehara, Y., 2014. Biofene, a renewable monomer for elastomer materials with novel properties: Polymer development, characterization and use. *Rubber World*, 25–30.

Scholz, T., Demharter, W., Hensel, R., and Kandler, O., 1987. Bacillus pallidus sp. nov., a new thermophilic species from sewage. *Systematic and Applied Microbiology*, 9, 91–96.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic Acids Research*, 33, W382–W388.

Seemann, M., Zhai, G., de Kraker, J.-W., Paschall, C.M., Christianson, D.W., and Cane, D.E., 2002. Pentalenene Synthase. Analysis of Active Site Residues by Site-Directed Mutagenesis. *Journal of the American Chemical Society*, 124, 7681–7689.

Sekiguchi, Y., Muramatsu, M., Imachi, H., Narihiro, T., Ohashi, A., Harada, H., Hanada, S., and Kamagata, Y., 2008. Thermodesulfovibrio aggregans sp. nov. and Thermodesulfovibrio thiophilus sp. nov., anaerobic, thermophilic, sulfate-reducing bacteria isolated from thermophilic methanogenic sludge, and emended description of the genus Thermodesulfovibrio. *International Journal of Systematic and Evolutionary Microbiology*, 58(11), 2541–2548.

Seo, S., Gomi, K., Kaku, H., Abe, H., Seto, H., Nakatsu, S., Neya, M., Kobayashi, M., Nakaho, K., Ichinose, Y., Mitsuhara, I., and Ohashi, Y., 2012. Identification of Natural Diterpenes that Inhibit Bacterial Wilt Disease in Tobacco, Tomato and Arabidopsis. *Plant Cell Physiology*, 53(8), 1432–1444.

Shaji Kumar, M.D., Abdulla Bava, K., Michael Gromiha, M., Prabakaran, P., Kitajima, K., Uedaira, H., and Sarai, A., 2006. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Research*, 34, D204–D206.

Sharkey, T.D. and Monson, R.K., 2017. Isoprene research – 60 years later, the biology is still enigmatic. *Plant Cell and Environment*, 40(9), 1671–1678.

Sharkey, T.D., Wiberley, A.E., and Donohue, A.R., 2008. Isoprene Emission from Plants: Why and How. *Annals of Botany*, 101(1), 5–18.

Shaw, J.J., Berbasova, T., Sasaki, T., Jefferson-George, K., Spakowicz, D.J., Dunican, B.F., Portero, C.E., Narváez-Trujillo, A., and Strobel, S.A., 2015. Identification of a Fungal 1,8-Cineole Synthase from Hypoxylon sp. with Specificity Determinants in Common with the Plant Synthases. *The Journal of Biological Chemistry*, 290(13), 8511–8526.

Shemesh, C., Pasvolsky, M., Sela, R., Green, N.J., and Zakin, S.J., 2013. Draft Genome Sequence of Alicyclobacillus acidoterrestris Strain ATCC 49025. *Genome Announcements*, 1(5).

Shields, V.D.C. and Hildebrand, J.G., 2001. Responses of a population of antennal olfactory receptor cells in the female moth Manduca sexta to plant-associated volatile organic compounds. *Journal of Comparative Physiology A*, 186(12), 1135–1151.

Shintani, M., Ohtsubo, Y., Fukuda, K., Hosoyama, A., Ohji, S., Yamazoe, A., Fujita, N., Nagata, Y., Tsuda, M., Hatta, T., and Kimbara, K., 2014. Complete genome sequence of the thermophilic polychlorinated biphenyl degrader Geobacillus sp.

strain JF8 (NBRC 109937). *Genome Announcements*, 2(1), 1213–1226.

Show, P.L., Oladele, K.O., Siew, Q.Y., Zakry, F.A.A., Lan, J.C.-W., and Ling, T.C., 2015. Overview of citric acid production from Aspergillus niger. *Frontiers in Life Science*, 8(3), 271–283.

Siedenburg, G. and Jendrossek, D., 2011. Squalene-Hopene Cyclases. *Applied and Environmental Microbiology*, 77(12), 3905–3915.

Sigrist, C.J.A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., and Bucher, P., 2002. PROSITE: A documented database using patterns and profiles as motif descriptors. *Briefings in Bioinformatics*, 3(3), 265–274.

Siliakus, M.F., Van Der Oost, J., and Kengen, S.W.M., 2017. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles*, 21(4), 651–670.

Simbahan, J., Drijber, R., and Blum, P., 2004. Alicyclobacillus vulcanalis sp. nov., a thermophilic acidophilic bacterium isolated from Coso Hot Springs, California, USA. *International Journal of Systematic and Evolutionary Microbiology*, 54, 1703–1707.

Singh, N.K., Carlson, C., Sani, R.K., and Venkateswaran, K., 2017. Draft Genome Sequences of Thermophiles Isolated from Yates Shaft, a Deep-Subsurface Environment. *Genome Announcements*, 5(22), 12–13.

Singh, R., Kumar, M., Mittal, A., and Mehta, P.K., 2017. Microbial metabolites in nutrition, healthcare and agriculture. *3 Biotech*, 7(1), 1–14.

Skewes-Cox, P., Sharpton, T.J., Pollard, K.S., and Derisi, J.L., 2014. Profile Hidden Markov Models for the Detection of Viruses within Metagenomic Sequence Data. *PLoS ONE*, 9(8), 1–12.

Slobodkina, G.B., Kovaleva, O.L., Miroshnichenko, M.L., Slobodkin, A.I., Kolganova, T. V, Novikov, A.A., van Heerden, E., and Bonch-Osmolovskaya, E.A., 2015. Thermogutta terrifontis gen. nov., sp. nov. and Thermogutta hypogea sp. nov., thermophilic anaerobic representatives of the phylum Planctomycetes. *International Journal of Systematic and Evolutionary Microbiology*, 65, 760–765.

Smanski, M.J., Peterson, R.M., Huang, S., and Shen, B., 2012. Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Current Opinion in Chemical Biology*, 16(1–2), 132–141.

Srividya, N., Lange, I., and Lange, B.M., 2016. Generation and Functional Evaluation of Designer Monoterpene Synthases. *In*: *Methods in Enzymology, Volume 576*. 147–165.

Starks, C.M., Back, K., Chappell, J., and Noel, J.P., 1997. Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase. *Science*, 277(5333), 1815–1820.

Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R., 1998. Sesquiterpene Synthases from Grand Fir (Abies grandis): Comparison of Constitutive and wound-induced activities, and cDNA Isolation, Characterization, and Bacterial Expression of δ-Selinene Synthase and γ-Humulene Synthase. *The Journal of Biological Chemistry*, 273(4), 2078–2089.

Stenhuus, B., Smits, H.P., Durhuus, T., and Katz, M., 2009. Production of stilbenoids. *Patent WO 2009124879 A2*.

Stepankova, V., Bidmanova, S., Koudelakova, T., Prokop, Z., Chaloupkova, R., and Damborsky, J., 2013. Strategies for Stabilization of Enzymes in Organic Solvents. *ACS Catalysis*, 3(12), 2823–2836.

Stephens, D.E., Singh, S., and Permaul, K., 2009. Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability. *FEMS Microbiology Letters*, 293(1), 42–47.

Stolyar, C., Liu, S., Thiel, Z., Tomsho, V.P., Pinel, L.P., Nelson, N.C., Lindemann, W.C., Romine, S.R., Haruta, M.F., Schuster, S.C., Bryant, S.C., and Fredrickson, D.A., 2014. Genome sequence of the thermophilic cyanobacterium Thermosynechococcus sp. strain NK55a. *Genome Announcements*, 2(1), 1060–1073.

Styles, M.Q., Nesbitt, E.A., Marr, S., Hutchby, M., and Leak, D.J., 2017. Characterization of the first naturally thermostable terpene synthases and development of strategies to improve thermostability in this family of enzymes. *The FEBS Journal*, 284(11), 1700–1711.

Suhara, Y., Wada, A., Tachibana, Y., Watanabe, M., Nakamura, K., Nakagawa, K., and Okano, T., 2010. Structure-activity relationships in the conversion of vitamin K analogues into menaquinone-4. Substrates essential to the synthesis of menaquinone-4 in cultured human cell lines. *Bioorganic & Medicinal Chemistry*, 18, 3116–3124.

Sultan, I., Rahman, S., Jan, A.T., Siddiqui, M.T., Mondal, A.H., and Haq, Q.M.R., 2018. Antibiotics, Resistome and Resistance Mechanisms: A Bacterial Perspective. *Frontiers in Microbiology*, 9, 1–16.

Sun, X., Shen, X., Jain, R., Lin, Y., Wang, J., Sun, J., Wang, J., Yan, Y., and Yuan, Q., 2015. Synthesis of chemicals by metabolic engineering of microbes. *Chemical Society Reviews*, 44(11), 3760–3785.

Sun, Y., Hong, H., Samboskyy, M., Mironenko, T., Leadlay, P.F., and Haydock, S.F., 2006. Organization of the biosynthetic gene cluster in Streptomyces sp. DSM 4137 for the novel neuroprotectant polyketide meridamycin. *Microbiology*, 152, 3507–3515.

Süssmuth, R.D. and Mainz, A., 2017. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition*, 56(14), 3770–3821.

Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M., and Abe, S., 1983. Bacillus thermoglucosidasius sp. nov., a new species of obligately thermophilic bacilli. *Systematic and Applied Microbiology*, 4(4), 487–495.

Takano, K., Higashi, R., Okada, J., Mukaiyama, A., Tadokoro, T., Koga, Y., and Kanaya, S., 2009. Proline Effect on the Thermostability and Slow Unfolding of a Hyperthermophilic Protein. *The Journal of Biochemistry*, 145(1), 79–85.

Tan, H.-T., Corbin, K.R., and Fincher, G.B., 2016. Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls. *Frontiers in Plant Science*, 7(December), 1–18.

Tan, Y. and Ji, G., 2010. Bacterial community structure and dominant bacteria in activated sludge from a 70 °C ultrasound-enhanced anaerobic reactor for treating carbazole-containing wastewater. *Bioresource Technology*, 101(1), 174–180.

Tang, Y.J., Sapra, R., Joyner, D., C, H.T., Myers, S., Reichmuth, D., Blanch, H., and Keasling, J.D., 2009. Analysis of Metabolic Pathways and Fluxes in a Newly Discovered Thermophilic and Ethanol-Tolerant Geobacillus Strain. *Biotechnology and Bioengineering*, 102(5), 1377–1386.

Tank, M. and Bryant, D.A., 2015. Chloracidobacterium thermophilum gen. nov., sp. nov.: An anoxygenic microaerophilic chlorophotoheterotrophic acidobacterium. *International Journal of Systematic and Evolutionary Microbiology*, 65(5), 1426–1430.

Tantillo, D.J., 2011. Biosynthesis via carbocations: Theoretical studies on terpene formation. *Natural Product Reports*, 28(6), 1035–1053.

Tholl, D., 2015. Biosynthesis and Biological Functions of Terpenoids in Plants. *In*: *Biotechnology of Isoprenoids*. 63–106.

Tian, B.-X., Wallrapp, F.H., Holiday, G.L., Chow, J.-Y., Babbitt, P.C., Poulter, C.D., and Jacobson, M.P., 2014. Predicting the Functions and Specificity of Triterpenoid Synthases: A Mechanism-Based Multi-intermediate Docking Approach. *PLoS Computational Biology*, 10(10), 1–16.

Tian, J., Wang, P., Gao, S., Chu, X., Wu, N., and Fan, Y., 2010. Enhanced thermostability of methyl parathion hydrolase from Ochrobactrum sp. M231 by rational engineering of a glycine to proline mutation. *The FEBS Journal*, 277(23), 4901–4908.

Tian, J., Wu, N., Chu, X., and Fan, Y., 2010. Predicting changes in protein thermostability brought about by single-or multi-site mutations. *BMC Bioinformatics*, 11, 370.

Torpenholt, S., de Maria, L., Olsson, M.H.M., Christensen, L.H., Skjøt, M., Westh, P., Jensen, J.H., and Lo Leggio, L., 2015. Effect of mutations on the thermostability of Aspergillus aculeatus β-1,4-galactanase. *Computational and Structural Biotechnology Journal*.

Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., Kato, N., and Sassa, T., 2007. Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *PNAS*, 104(9), 3084–3088.

Trapp, S. and Croteau, R., 2001a. Defensive Resin Biosynthesis in Conifers. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52, 689–724.

Trapp, S.C. and Croteau, R.B., 2001b. Genomic Organization of Plant Terpene Synthases and Molecular Evolutionary Implications. *Genetics*, 158(2), 811–832.

Trott, O. and Olson, A.J., 2010. Autodock vina: improving the speed and accuracy of docking. *Journal of Computational Chemistry*, 31(2), 455–461.

Turner, P., Mamo, G., and Karlsson, E.N., 2007. Microbial Cell Factories Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories*, 6, 1–23.

Tyc, O., Song, C., Dickschat, J.S., Vos, M., and Garbeva, P., 2016. The Ecological

Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends in Microbiology*, 25(4), 280–292.

Vannice, J.C., Skaff, D.A., Keightley, A., Addo, J.K., Wyckoff, G.J., and Miziorko, H.M., 2014. Identification in Haloferax volcanii of Phosphomevalonate Decarboxylase and Isopentenyl Phosphate Kinase as Catalysts of the Terminal Enzyme Reactions in an Archaeal Alternate Mevalonate Pathway. *Journal of Bacteriology*, 196(5), 1055–1063.

Vardakou, M., Salmon, M., Faraldos, J.A., and O'Maille, P.E., 2014. Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases. *MethodsX*, 1, 187–196.

Vattekkatte, A., Garms, S., Brandt, W., and Boland, W., 2018. Enhanced structural diversity in terpenoid biosynthesis: enzymes, substrates and cofactors. *Organic & Biomolecular Chemistry*, 16, 348.

Veith, A., Botelho, H.M., Kindinger, F., Gomes, C.M., and Kletzin, A., 2012. The Sulfur Oxygenase Reductase from the Mesophilic Bacterium Halothiobacillus neapolitanus Is a Highly Active Thermozyme. *Journal of Bacteriology*, 194(3), 677–685.

Veno, J., Kamarudin, N.H.A., Shukuri, M., Ali, M.M., Masomian, M., and Rahman, R.N.Z.R.A., 2017. Directed Evolution of Recombinant C-Terminal Truncated Staphylococcus epidermidis Lipase AT2 for the Enhancement of Thermostability. *International Journal of Molecular Sciences*, 18.

Vickers, C.E., Gershenzon, J., Lerdau, M.T., and Loreto, F., 2009. A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nature Chemical Biology*, 5(5), 283–291.

Vieille, C. and Zeikus, G.J., 2001. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiology and Molecular Biology Reviews*, 65(1), 1–43.

Vinokur, J.M., Korman, T.P., Cao, Z., and Bowie, J.U., 2014. Evidence of a Novel Mevalonate Pathway in Archaea. *Biochemistry*, 53(25), 4161–4168.

Wallach, O., 1887. Zur Kenntnis der Terpene und ätherischen Öle, vierte Abhandlung. *Justus Liebigs Annalen der Chemie*, 238(1–2), 78–88.

Wang, C., Liwei, M., Park, J. Bin, Jeong, S.H., Wei, G., Wang, Y., and Kim, S.-W.,

2018. Microbial platform for terpenoid production: Escherichia coli and Yeast. *Frontiers in Microbiology*, 9, 1–8.

Wang, C., Park, J.E., Choi, E.S., and Kim, S.W., 2016. Farnesol production in Escherichia coli through the construction of a farnesol biosynthesis pathway – application of PgpB and YbjG phosphatases. *Biotechnology Journal*, 11(10), 1291–1297.

Wang, G., Tang, W., and Bidigare, R.R., 2005. Terpenoids As Therapeutic Drugs and Pharmaceutical Agents. *In*: *Natural Products: Drug Discovery and Therapeutic Medicine*. 197–227.

Wang, J., Guleria, S., Koffas, M.A.G., and Yan, Y., 2016. Microbial production of value-added nutraceuticals. *Current Opinion in Biotechnology*, 37, 97–104.

Wang, J., Kodali, S., Lee, S.H., Galgoci, A., Painter, R., Dorso, K., Racine, F., Motyl, M., Hernandez, L., Tinney, E., Colletti, S.L., Herath, K., Cummings, R., Salazar, O., González, I., Basilio, A., Vicente, F., Genilloud, O., Pelaez, F., Jayasuriya, H., Young, K., Cully, D.F., and Singh, S.B., 2007. Discovery of platencin, a dual FabF and FabH inhibitor with in vivo antibiotic properties. *PNAS*, 104(18), 7612–7616.

Wang, J., Soisson, S.M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Painter, R., Parthasarathy, G., Tang, Y.S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya, H., Ondeyka, J., Herath, K., Zhang, C., Hernandez, L., Allocco, J., Basilio, Á., Tormo, J.R., Genilloud, O., Vicente, F., Pelaez, F., Colwell, L., Ho Lee, S., Michael, B., Felcetto, T., Gill, C., Silver, L.L., Hermes, J.D., Bartizal, K., Barrett, J., Schmatz, D., Becker, J.W., Cully, D., and Singh, S.B., 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature*, 441(7091), 358–361.

Wang, P., Li, L., Chen, X., Jiang, N., Liu, G., Chen, L., Xu, J., Song, H., Chen, Z., and Ma, Y., 2012. Draft Genome Sequence of Alicyclobacillus hesperidum Strain URH17-3-68. *Journal of Bacteriology*, 194(22), 6348.

Wang, Q., Quan, S., and Xiao, H., 2019. Towards efficient terpenoid biosynthesis: manipulating IPP and DMAPP supply. *Bioresources and Bioprocessing*, 6(6), 1–13.

Wang, Z., Yeats, T., Han, H., and Jetter, R., 2010. Cloning and Characterization of Oxidosqualene Cyclases from Kalanchoe daigremontiana. *The Journal of*

Biological Chemistry, 285(39), 29703–29712.

Warth, A.D., 1978. Relationship Between the Heat Resistance of Spores and the Optimum and Maximum Growth Temperatures of Bacillus Species. *Journal of Bacteriology*, 134(3), 699–705.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46, W296–W303.

Weathers, P.J., Arsenault, P.R., Covello, P.S., McMickle, A., Teoh, K.H., and Reed, D.W., 2011. Artemisinin production in Artemisia annua: studies in planta and results of a novel delivery method for treating malaria and other neglected diseases. *Phytochemistry Reviews*, 10(2), 173–183.

Wei, G., Jia, Q., Chen, X., Köllner, T.G., Bhattacharya, D., Wong, G.K.-S., Gershenzon, J., and Chen, F., 2019. Terpene Biosynthesis in Red Algae Is Catalyzed by Microbial Type But Not Typical Plant Terpene Synthases. *Plant Physiology*, 179(2), 382–390.

Wei, J.H., Yin, X., and Welander, P. V, 2016. Sterol synthesis in bacteria. *Frontiers in Microbiology*, 7.

Weitman, M. and Major, D.T., 2010. Challenges Posed to Bornyl Diphosphate Synthase: Diverging Reaction Mechanisms in Monoterpenes. *Journal of the American Chemical Society*, 132(18), 6349–6360.

Wendt, K.U., Poralla, K., and Schulz, G.E., 1997. Structure and Function of a Squalene Cyclase. *Science*, 277(5333), 1811–1815.

Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., Fickes, S., Diola, D., Benjamin, K.R., Keasling, J.D., Leavell, M.D., Mcphee, D.J., Renninger, N.S., Newman, J.D., and Paddon, C.J., 2012. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *PNAS*, 109(3), E111–E118.

Weston, L.A. and Mathesius, U., 2013. Flavonoids: Their Structure, Biosynthesis and Role in the Rhizosphere, Including Allelopathy. *Journal of Chemical Ecology*, 39(2), 283–297.

Whited, G.M., Feher, F.J., Benko, D.A., Cervin, M.A., Chotani, G.K., McAuliffe, J.C., La Duca, R.J., Ben-Shoshan, E.A., and Sanford, K.J., 2010. Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering. *Industrial Biotechnology*, 6(3), 152–163.

Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R.B., and Christianson, D.W., 2002. Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *PNAS*, 99(24), 15375–15380.

Wilding, E.I., Brown, J.R., Bryant, A.P., Chalker, A.F., Holmes, D.J., Ingraham, K.A., Iordanescu, S., So, C.Y., Rosenberg, M., and Gwynn, M.N., 2000. Identification, Evolution, and Essentiality of the Mevalonate Pathway for Isopentenyl Diphosphate Biosynthesis in Gram-Positive Cocci. *Journal of Bacteriology*, 182(15), 4319–4327.

Woodside, A.B., Huang, Z., and Poulter, C.D., 1993. TRISAMMONIUM GERANYL DIPHOSPHATE. *Organic Syntheses Collective*, 8, 616–621.

Wu, H., Liu, B., Shao, Y., Ou, X., and Huang, F., 2018. Thermostaphylospora grisealba gen. nov., sp. nov., isolated from mushroom compost and transfer of Thermomonospora chromogena Zhang et al. 1998 to Thermostaphylospora chromogena comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 68(2), 602–608.

Xu, M., Hillwig, M.L., Lane, A.L., Tiernan, M.S., Moore, B.S., and Peters, R.J., 2014. Characterization of an Orphan Diterpenoid Biosynthetic Operon from Salinispora arenicola. *Journal of Natural Products*, 77(9), 2144–2147.

Xu, M., Wilderman, P.R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N.M., Coates, R.M., and Peters, R.J., 2006. Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry*, 68(3), 312–326.

Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M., and Yokota, A., 2011. Thermasporomyces composti gen. nov., sp. nov., a thermophilic actinomycete isolated from compost. *International Journal of Systematic and Evolutionary Microbiology*, 61, 86–90.

Yamada, Y., Arima, S., Nagamitsu, T., Johmoto, K., Uekusa, H., Eguchi, T., Cane, D.E., Ikeda, H., Shin-ya, K., Cane, D.E., and Ikeda, H., 2016. Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in

an engineered Streptomyces host. *The Journal of Antibiotics*, 68(6), 385–394.

Yamada, Y., Cane, D.E., and Ikeda, H., 2012. Diversity and analysis of bacterial terpene synthases. *In*: *Methods in Enzymology*. 123–162.

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-Ya, K., Omura, S., Cane, D.E., and Ikeda, H., 2015. Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 112(3), 857–862.

Yamaguchi, S., 2008. Gibberellin Metabolism and its Regulation. *Annual Review of Plant Biology*, 59, 225–251.

Yamashita, S., Hemmi, H., Ikeda, Y., Nakayama, T., and Nishino, T., 2004. Type 2 isopentenyl diphosphate isomerase from a thermoacidophilic archaeon Sulfolobus shibatae. *European Journal of Biochemistry*, 271(6), 1087–1093.

Yáñez-Serrano, A.M., Fasbender, L., Kreuzwieser, J., Dubbert, D., Haberstroh, S., Lobo-Do-Vale, R., Caldeira, M.C., and Werner, C., 2018. Volatile diterpene emission by two Mediterranean Cistaceae shrubs. *Scientific Reports*, 8, 1–13.

Yang, C., Chen, H., Chen, H., Zhong, B., Luo, X., and Chun, J., 2017. Antioxidant and anticancer activities of essential oil from gannan navel orange peel. *Molecules*, 22(8), 1–10.

Yang, J. and Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, 43, W174–W181.

Yang, L., Wen, K.-S., Ruan, X., Zhao, Y.-X., Wei, F., and Wang, Q., 2018. Response of Plant Secondary Metabolites to Environmental Factors. *Molecules*, 23(4), 1–26.

Yang, S., Tian, H., Sun, B., Liu, Y., Hao, Y., and Lv, Y., 2016. One-pot synthesis of (−)-Ambrox. *Scientific Reports*, 6, 1–6.

Yasawong, M., Areekit, S., Pakpitchareon, A., Santiwatanakul, S., and Chansiri, K., 2011. Characterization of Thermophilic Halotolerant Aeribacillus pallidus TD1 from Tao Dam Hot Spring, Thailand. *International Journal of Molecular Sciences*, 12(8), 5294–5303.

Ye, L., Lv, X., and Yu, H., 2016. Engineering microbes for isoprene production. *Metabolic Engineering*, 38, 125–138.

Yi, Z.-L., Pei, X.-Q., and Wu, Z.-L., 2010. Introduction of glycine and proline residues

onto protein surface increases the thermostability of endoglucanase CelA from Clostridium thermocellum. *Bioresource Technology*, 102(3), 3636–3638.

Yoon, B., 2009. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6), 402–415.

Yoon, J.H., Kim, I.G., Shin, Y.K., and Park, Y.H., 2005. Proposal of the genus Thermoactinomyces sensu stricto and three new genera, Laceyella, Thermoflavimicrobium and Seinonella, on the basis of phenotypic, phylogenetic and chemotaxonomic analyses. *International Journal of Systematic and Evolutionary Microbiology*, 55(1), 395–400.

Yoshikuni, Y., Ferrin, T.E., and Keasling, J.D., 2006. Designed divergent evolution of enzyme function. *Nature*, 440(7087), 1078–1082.

Yoshikuni, Y., Martin, V.J.J., Ferrin, T.E., and Keasling, J.D., 2006. Engineering Cotton (+)-δ-Cadinene Synthase to an Altered Function: Germacrene D-4-ol Synthase. *Chemistry & Biology*, 13(1), 91–98.

Yoshinaka, T., Yano, K., and Yamaguchi, H., 1973. Isolation of Highly Radioresistant Bacterium, Arthrobacter radiotolerans nov. sp. *Agricultural and Biological Chemistry*, 37(10), 2269–2275.

You, S., Yin, Q., Zhang, J., Zhang, C., Qi, W., Gao, L., Tao, Z., Su, R., and He, Z., 2017. Utilization of biodiesel by-product as substrate for high-production of β-farnesene via relatively balanced mevalonate pathway in Escherichia coli. *Bioresource Technology*, 243, 228–236.

Yu, H. and Huang, H., 2013. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnology Advances*, 32, 308–315.

Yu, T.T., Zhang, B.H., Yao, J.C., Tang, S.K., Zhou, E.M., Yin, Y.R., Wei, D.Q., Ming, H., and Li, W.J., 2012. Lihuaxuella thermophila gen. nov., sp. nov., isolated from a geothermal soil sample in Tengchong, Yunnan, south-west China. *Antonie van Leeuwenhoek*, 102(4), 711–718.

Yu, X., Lin, H., Wang, Y., Lv, W., Zhang, S., Qian, Y., Deng, X., Feng, N., Yu, H., and Qian, B., 2018. D-limonene exhibits antitumor activity by inducing autophagy and apoptosis in lung cancer. *OncoTargets and Therapy*, 11, 1833–1847.

Yu, X., Zhang, Y., Ma, Y., Xu, Z., Wang, G., and Xia, L., 2013. Expression of an (E)-β-farnesene synthase gene from Asian peppermint in tobacco affected aphid

infestation. *The Crop Journal*, 1(1), 50–60.

Zada, B., Wang, C., Park, J.-B., Jeong, S.-H., Park, J.-E., Birla Singh, H., and Kim, S.-W., 2018. Metabolic engineering of Escherichia coli for production of mixed isoprenoid alcohols and their derivatives. *Biotechnology for Biofuels*, 11, 210.

Zebec, Z., Wilkes, J., Jervis, A.J., Scrutton, N.S., Takano, E., and Breitling, R., 2016. Towards synthesis of monoterpenes and derivatives using synthetic biology. *Current Opinion in Chemical Biology*, 34, 37–43.

Zeigler, D.R., 2014. The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology*, 160, 1–11.

Zeldes, B.M., Keller, M.W., Loder, A.J., Straub, C.T., Adams, M.W.W., and Kelly, R.M., 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology*, 6, 1–17.

Zhang, Y., Taiming, L., and Liu, J., 2003. Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in Escherichia coli. *Protein Expression and Purification*, 29(1), 132–139.

Zhao, B., Lin, X., Lei, L., Lamb, D.C., Kelly, S.L., Waterman, M.R., and Cane, D.E., 2008. Biosynthesis of the Sesquiterpene Antibiotic Albaflavenone in Streptomyces coelicolor A3(2). *The Journal of Biological Chemistry*, 283(13), 8183–8189.

Zhao, J., Lou, J., Mou, Y., Li, P., Wu, J., and Zhou, L., 2011. Diterpenoid Tanshinones and Phenolic Acids from Cultured Hairy Roots of Salvia miltiorrhiza Bunge and Their Antimicrobial Activities. *Molecules*, 16(3), 2259–2267.

Zhen-Li, Z., Xin-Qi, Z., Nan, W., Wen-Wu, Z., Xu-Fen, Z., Yi, C., Min, W., and Wu, M., 2014. Amphiplicatus metriothermophilus gen. nov., sp. nov., a thermotolerant alphaproteobacterium isolated from a hot spring. *International Journal of Systematic and Evolutionary Microbiology*, 64(8), 2805–2811.

Zheng, Y., Liu, Q., Li, L., Qin, W., Yang, J., Zhang, H., Jiang, X., Cheng, T., Liu, W., Xu, X., and Xian, M., 2013. Metabolic engineering of Escherichia coli for high-specificity production of isoprenol and prenol as next generation of biofuels. *Biotechnology for Biofuels*, 6, 1–13.

Zhou, H., Chen, L., Liu, Y., Chen, J., and Francis, F., 2016. Use of slow-release plant infochemicals to control aphids: a first investigation in a Belgian wheat field. *Scientific Reports*, 6, 1–8.

Zhou, J., Wang, C., Yoon, S.-H., Jang, H.-J., Choi, E.-S., and Kim, S.-W., 2014. Engineering Escherichia coli for selective geraniol production with minimized endogenous dehydrogenation. *Journal of Biotechnology*, 169, 42–50.

Zhou, K., Gao, Y., Hoy, J.A., Mann, F.M., Honzatko, R.B., and Peters, R.J., 2012. Insights into Diterpene Cyclization from Structure of Bifunctional Abietadiene Synthase from Abies grandis. *Journal of Biological Chemistry*, 287(9), 6840–6850.

Zhou, K. and Peters, R.J., 2009. Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. *Phytochemistry*, 70, 366–369.

Zhou, Y., Gao, S., Wei, D.-Q., Yang, L.-L., Huang, X., He, J., Zhang, Y.-J., Tang, S.-K., and Li, W.-J., 2012. Paenibacillus thermophilus sp. nov., a novel bacterium isolated from a sediment of hot spring in Fujian province, China. *Antonie van Leeuwenhoek*, 102, 601–609.

Zyryanov, A.B., Shestakov, A.S., Lahti, R., and Baykov, A.A., 2002. Mechanism by which metal cofactors control substrate specificity in pyrophosphatase. *Biochemical Journal*, 367(3), 901–906.

8 Appendix

8.1 Codon optimised and codon harmonized genes for *E. coli* and *P. thermoglucosidasius*

The codon optimised DNA sequence for *E. coli* of *tcur_3107* from *T. curvata* DSM 43183:

AAAAAGAGCTCAGGAGGGCTAGCATGGATGCAGCACTGGTTCTGAGCCTGGCAGATCGTGTTGC
AAGTCTGGCCGATCGTAGCGGTATGCATCCGGCAGCACAGCTGATTGGTGCACAGGCAGAAGGT
TGGGCACGTAGCCGTGGTCTGCTGCTGGGTGATCCGGATGCAACACCGCTGGGTCGTGCACGTT
TTGAACGTCTGGCATGTCGTATTTTTCCGCATGCACAGCCGGATCGTGTGGTTCTGTTTGCACGTT
GGCTGATGTGGCTGTTTGCCCTGGATGATCATTTTGATGATACTCCGCTGGGTGCAAGCGCAACC
AGCGTTGATGGTCTGTATGCCGATCTGCTGGGAGCACTGCGTCGTGGTCATACCAAACCGGAAG
CCGGTGCACTGGAACTGGCACTGGAAGAACTGTGGCGTGATACCGTTCCGGGTACAAGTCCGCA
GTGGCGTCATCATTTTCTGCGTCTGATGGAAGAACATCGTGCAGCATGTGCAGAAGAAGCAGTTA
ATCGTCGTACCGGTCGTATTGCACCGCTGGCAGATTATCCGGTTCTGCGTCGTCGTAGCGCAGG
TCCGTTTCTGTATGAACTGGCCGAACCTGTTCTGCAGGTTGCACTTGATCCGCGTCTGAAACGTA
GTCCGGCATGGAAAGCCCTGGTTGATGGCACCGCAGATATGATTACCTGGGCAAATGATGTTGTT
AGCTATCCGAAAGAAAGCCGTCAGGGTACAGTTCCGGTGACCGGTAATCTGGTTGCAGTTGCATG
TCGTGAACTGGGTATGGCACCAGCCCAGGCAGCAAGCTGGGTTGTTGATCGTATTGCCCGTCGT
GCACCGCAGGTTCGTGAAGCAGCACGTGCAGTTGGTGCAGAACTGGATCGTCTGGAAATTGGTC
CGCAGGGTCGTAAAGATACCGCAGCAGTTGTTCGTGTTCTGCTGCAGGCACCGCGTGCACACAT
GGATTGGCTGGCAGAAACCGGTCGTTATACCCCTCCGGTTCGTTCACCGGTTGTGCTGCTGCATC
GTACCGCAGCCGGTGTTGCCCGTACCATTGGTTAAGAATTCAAAAA

The codon optimised DNA sequence for *E. coli* of *fjsc11* from *F. thermalis*:

AAAAAGAGCTCAGGAGGGCTAGCATGAACGAATTTACCCTGCCGGAACTGTATTGTCCGTTTCCG
TATCAGATTAACAATTATGCCGATGTGCTGGAAAAACATGCACTGGAATGGGTTCTGAAATATAAT
CTGCTGATCGATGAGAGCTACTACCAGTATTTTTGCAAGAGCAAAATCTTCCTGCTGATTGCAGGT
AGCTATCCGTATTGTGATCTGGAAGAACTGAAAGTTGCAAATGATTGGCTGACCTGGATGATCTTT
CTGGATGATTATTATGATACCAGCGACTTCAAAAACAAGCCCGAACTGATTACCAAACTGCACAAT
CGCTTTTTTGAAATTCTGAGCGGTGCCGAAATTACCAATCAGGATACCCCGTATAGCCATGCACTG
AATAATCTGCGTCATCGTACCCTGCAGATTGGCAATCCGCGTTGGTTTCATTTTTTTGTTTGTGCC
CTGAGCGAGTTCCTGGATGGTTGTGTTCAAGAAGCACATAATCGCGCAAATGGTATTCTGCCGGA
TATTGAAACCTATATTATCCTGCGTCGTCTGACCGGTGGTATGGGTCCGCTGTTTGAACTGATTGA
ATTTTGCAACCATCTGGAAATTCCGTATCTGCTGCGTGAAAACATCATCCTGAAAAAGCTGAAAAT
GATGAGCAACAACATTATCTGCTGGTGCAACGATATCTATAGCATTCCGAAAGAACTGCGTATTGG
TGATCCGCATAATCTGGTGCTGTTACTGAAAAACCATAAAAAGATCAGCCTGAAACATGCGATTAC
CCAGGTTAGCGAAATGCATGATCAAGAGGTTCAGCGTATGATTGAACTGGAAAGCACCCTGCCGT
GTCTGGGTCAAGAACTGGATGCAGAACTGGCCAAATATATCAGCGGTATTCATGCCTGGATTGCC
AGCCATTTTCATTGGTATAGTCATAGCGGTCGTTATGAAGTTACCGAAAAACTGGCCCTGGAAGAA
GATGTTAAACTGGTTAATGCCTAAGAATTCAAAAA

The codon optimised DNA sequence for *E. coli* of *tchrom* from *T. chromogena*:

AAAAAGAGCTCAGGAGGGCTAGCATGGAAGAAGTTGTTCTGCCGGAACTGCATATTCCGTTTCCG
AGCCGTCTGAATCCGTTTGAAGCAGAAGCAGCAGCACGTACCACCGCATGGGCACGTGAATTTG
GTCTGATTAAAGAAGAAAAAGAAGCAGCCCGTTTTGCCGGTGAAAGCTATGCACGTCTGGTTGCC
CGTCTGTATCCGGAAGCAAGCCTGACCGATCTGGTTCTGGCAGCAGATCTGAATAGCTGGTTTCA
TGTTTTTGATGATCAGTTTGAACTGGCCGAAGTTGGTCGTGATCCGGAACTGGCACGTCGTCTGG
CAGAACATGCAGATAGCCTGATGAAAGGTGAACCGCTGACCAGCCGTGCAGGTCCGGTTCTGAC
CGCACTGGCCGATCTGCGTGATCGTCTGCGTTATCGTGCCGGTGATACCTGGTGGGAACGTTTT
GCAGATCATATGCGTCAGTGTCTGGATGCAGCACTGTGGGAAGTTGATAATCGTGCACGTGAAAC
CGTTCCAGATCCGGTTACCTATGTTGATCGTAAACTGCTGATTGCATATGTTCCGCCTAGCTTTGA
TATTATCGAACTGGTTGAACATGTGGAACTGCCGAATGCAATTCGTCATAGTCCGGAATATCAGAC
CCTGCTGCATGAAGCCGGTCATGTTGTTGTTTGTACCAATGATGTTGTTGGTCTGCGTCGTGAACT
GCTGCAGGGTGAATTTCATAATCTGGTTATTGTTCTGCGTCATGCAATGGGTTGTACCCTGCAAGA
AGCAACGGATCAGGTTGCCACCACCATTGAAGAACGTGTTGGTCGCTATCTGGAAGCAAAAACAG
CACTGGAAGAACGTATGGATCGTCTGGGTGTTACCGGTCCGGAACGTAGCGCAGTTCGTCGTTG
TGTGACCGGTCTGGAAGATTGGATGCGTGGTTATCTGGATTGGGCATTAGAAACCCGTCGTTTTA
CCGATTTTGTTTATCGTGGTGAAGTTGCCAGCTTTCATGGTGAACTGGTGGATTAAGAATTCAAAA
A

The codon optimised DNA sequence for *E. coli* of *tbis_3257* from *T. bispora* DSM 43833:

AAAAAGAGCTCAGGAGGGCTAGCATGATTGCACGTAGCTTTGGTCCGGCAGAAGCAGAAGCCGC
AGGTCTGGGTCGTGCATGTGCAGTTGCAGTTCGTAGCGCACGTGAACTGATGGCAACCGCAGAA
GAATATCCGGAACTGTTTCCGCCTAAACCGTTTGATGCAGCATTTTTTAGCGGTCTGGCACTGAG
CAGCGCATTTGGTAGCCCGTGGGCAGGTCCGGAAGAACTGCGTGCAGTTAATCTGGCAAGCCTG
TGTGTTTTTGCAGTTGATTGGCAGGCAGAACGTGCAAGCACCGCAGGCGAAATGGATGGTCTGAT
TGCCCGTTGTCTGGCAGCAGCCGGTGGTGAAGCAGGCGGTACACCGATTGCACGCCTGATTGCA
GGTCTGCGTGAAGAACTGGCAGCCGCATGTCCGAGCTTTGGTGAACTGCAGCCGATTTGGGCAG
GCCAGCTGCATCGTATGCTGTTAGCAATGGCCACCGAACTGCGTTGGAAACGTGAAGCAGCACG
TGTTACCCTGGATGATTATCTGGCAAATGCAGATGGTTGTGGTGCAAGCTTTGTGAATATTTCACA
TTGGATTGCAACCGGTGATCCGTGGACACTGCGTAATCTGGCGGAACTGCGCGAAGTTAGCGCA
GAAGTTCAGCGTTATCTGCGTCTGCTGAATGATCTGGCAACCAGCGGTCGTGAACGTGAATGGG
GTGATCTGAGCGCACTGACCTTAGGTGCAGGTCGTGAAGAGGTTATTGAACGTATGGCAGGTATT
CTGGGTCGCTGTCGTGCACTGCTGGAACCGGTGCGTGCAGGTAGTCCGCGTGTTGCAGCATATC
TGGAACGTCAGATTGGTTTTAATACCGGTTTTTATGGTGTGGCCGATTATTGGGGTGAACTGTAAG
AATTCAAAAA

The codon optimised DNA sequence for *E. coli* of *jkg1* from *K. papyrolyticum* JKG1^{T}:

AAAAAGAGCTCAGGAGGGCTAGCATGGAACGTCTGCGTCTGCCGCAGCTGTTTTGTCCGTTTACC
ACCAGTTGTAGCCGTTATGTGGCAGTTGTTGAACGTGCAACCCTGGCATGGGCACGTCGTTATGG
TCTGATTACCAGCGTTGAAGCAGAACAGCGTCTGAGCGCAGTTCGTATTGGTGCACTGATTGGTC
GTGTTTATCCGATGCTGGCACGTCATGATCTGCAGCTGCTGGTTGATTGGACCACCTGGGGTTTT
ATTTGGGATGATCTGTGTAGCGCACCGCCTCTGCGTGATCAGCCGGAACAGCTGCAGCACCAGC

AGGCACGTCTGGCAGCAGTTCTGCGTGGTGCAGCACCGGAAGCAGATCAGCCGCTGGCAGGCG
CACTGGCCGATCTGCGTCGTCGTCTGCTGCAGAAAACCAGCGTGGTTGGTCTGGAACGTTTTGTT
CGTAGCGTTGAACAGTTTTTTGAAGCATGTCTGTGGGAAGCAACCAATCGTGCCCAGGGTCGTGT
TCCGGATCTGGCAAGCTATCAGCGTATGCGTCCGCGTAGCAGCGGCATTAATACCTATACCGAAC
TGTTTGGTATTATCGATGGTCTGGATTTTCCGGAAGCACTGCGTGCACATCCGACCGTTGAACGT
CTGACCCTGCTGGCAAATAATGTTGTTTGTTGGATCAACGATATTATCAGCCTGGCAAAAGAACTG
GAACAGGGTGATGTTCATAATCTGGTTCTGATCCTGCAGCATGAACAGCAGATTGGTCTGCAAGA
AGCACTGAATTGTACCGCAGCACTGGTTAATGCAGAAGTTCGTAGCTTTATTACCGCAACACGTTG
GCTGCCGGTTCTGGATGAACCGGGTACAGCAGCCCTGGAACGCTATCTGACCGTGCTGCGTGGC
TGGATGCGTGGTAGCCTGGAATGGTCATATACCAGCGGTCGTTATCTGAGCACCACACGTGTTAG
TCTGGATCTGCCGGAAACCATGTGTCTGGAATTCAAAAA

The codon harmonised DNA sequence for *E. coli* of *rxyl_0493* from Rubrobacter xylanophilus DSM 9941:

AAAAATCTAGAAATAATTTTGTTTAACTTTAAGGCTAGCAACAAAGAGGAGAGTGCAAAATCGATG
AGCACGGAAAACGGTCACCGTGATGGCCTGGGTCCGCTGCGTTGTCCGTTCCCGGCAGCAATCA
ACCCGCACGCGGATGAAGTACACCGTGAAACCGTAGAATGGGCGGAAGGCTTCGGTCTGCTGG
GTCCGGGTGATGGTCACCGTATGATGCGTGATACCGGTATCGGTCGTCTGGCGGGTCGTTTCCA
CCCGGGTGCGGGTCGTGAAGAACTGCGTCTGATCAGCGATTGGTGCGCGTGGATGTTCCTGCGT
GATGATCTGGCGGATGCACCGGCATACTTCCGTCACCCGGAACGTCTGGCAGCGCTGGATGCAG
CGTTCCTGGATATCCTGAGCGGTCGTGCGAGCGGTGAAGGTTGCGGTAGCTTCGGTCGTGCACT
GCGTGACCTGCGTGAACGTCTGCTGCCGAAAGTACCGGCACCGCTGTGGCTGCGTCGTTTCCTG
CGTAGCGTAGAAGAACACTTCGAAAGCACTCTGTGGGAAGCGACCAACCGTGCGCGTGGTGTAG
TACCGGATCTGGAAACCTACCTGCGTATGCGTCCGATCACCGGTGGTATGCACGTAGATACTGAT
TTCATCGAAATCAGCAGCGGTGTATACCTGCCGCCGGAAGTACGTCGTCACCCGGCGGTATCCG
CGCTGACCGGTGCGAGCAACAACGTAGTATGCTGGGCAAACGATATCATCTCCCTGGCGAAAGA
ACGTTCCCGTGGTGATGTACACAACCTGGTACTGGTACTGCGTGCATCCCGTCGCCTGACCACC
CGTGAAGCAGTAGCAGAAGCGGCGCGTATGTACGAAGCAGAAGTACGTCGTTTCGTACGTCTGG
AACGTGAACTGCCGCCGTTCGGTCCGGCGATCGATGCGAACCTGCGTCGTTACGTATCCGTACT
GAAATCCCGTATGCGTGGTAACCTGGATTGGACCTACGAATCCGCACGTTACCGTGCGGGTGCA
GCGAGCCGTTAAGAATTCAAAAA

The codon optimised DNA sequence for *E. coli* of *sscg_03688* from *S. clavuligerus* ATCC 27074:

AAAAAGCTAGCATGAGCCTGAATCATAGCGATCTGATGTTTTATTGTCCGGTTGATGATCTGCCGC
ATCCGGCAGCAAGCGGTGTTAATGATCGTACCCTGGATTGGGCAAGCGGTCAGGGTATTCCGAC
CGCAGATCGTGATGCAGGTCGTCTGCGTGCAATGGCACCGGGTCTGCTGGCAGCACGTATTGCA
CCGGATGCACGTGGTCCGGTTCTGGATGCATTTGCCGATCATCATACCTGGCTGTTTGCATTTGA
TGATGAATATTGTGATCGTGCAGATGGTAGCGGTATTACCGAATGGGCAAGCTTTCTGGCACGTC
TGCATCGTGTTGTTGAAACCGGTGAAAGCGCACTGCTGCCTGGTAATCCGTATGGTCTGGCACTG
CGTGATATTGCATGTCGTCTGAGCACCTATACCACACCGGCACAGCTGGCAGAATGGCTGGAAG
CACTGCGTAGCTATTTTGCAGCACTGGTTTGGGAACGTAGCCGTCGTCGTGATGATGACCGTCTG
CAGAGCCTGGATGATTATCTGCTGCTGCGTCTGCGTAATGGTGCAATGCATACCAGCATTACCCT
GCTGGATACCGTTAATGGTTATGTTCTGCCTCGTGAACTGCGTGAAACACCGGGTGTTCGTGCAC

TGGTTGAAATGACAGCACTGCTGGTTAGCGTTGATAATGATATTCTGAGCCATCACAAAGAAAGCA
CCAGCGGCACCCGTGAAGCAAATCTGCTGGATGTTCTGGGTCGTACCGGTCATACCACTCCGGG
TGAAGCAGTTGCACAGGCAGTTGCCCTGCGCAATGAAATTATGCGTCAGTTTGTTCGTGTTGCAG
AACGTGTTCGTACACCGGCAGCCGTGCCGGAACTGTATCGTTTTACCACAGGCCTGGCACGTTG
GATTCGTGCAAATCTGGATTTTTCACTGACCACCACCCGTTATACCGGTCCGGTTACCGAACGTG
CAGCCCTGAGTCCGCATGAAGTTCCGCCTCTGAGTGGTCAGGGTCCGGCACCGGCACGTAGTGA
TGTTATTGGTTGGTGGTGGCGTATTCCGGAACCGCTGCCTGAACCGGGTTCTGATGGTGCAGATA
CACCGGTTCGTAAACGTCGTGCGGGTGATCGTCCGCCTACCGCAGGTCGTGGTGGTGCACCGCA
TCATCAGCGCACCGGTCCGCCTCCGCCTGTTCTGCCAGGTGGCATTACCGCAAGCCGTAGCAGC
GGTCTGCAGCAGAGCACCTGGCGTCGTGAACATCGTTAAGAATTCAAAAA

The codon optimised DNA sequence for *E. coli* of *sds* from *S. pristinaespiralis*:

AAAAAGAGCTCAGGAGGGCTAGCATGGAACCGGAACTGACCGTTCCGCCTCTGTTTAGCCCGAT
TCGTCAGGCAATTCATCCGAAACATGCAGATATTGATGTTCAGACCGCAGCATGGGCAGAAACCT
TTCGTATTGGTAGCGAAGAACTGCGTGGTAAACTGGTTACCCAGGATATTGGCACCTTTAGCGCA
CGTATTCTGCCGGAAGGTCGTGAAGAAGTTGTTAGCCTGCTGGCAGATTTTATTCTGTGGCTGTTT
GGTGTTGATGATGGTCATTGTGAAGAAGGTGAACTGGGTCATCGTCCGGGTGATCTGGCAGGTC
TGCTGCATCGTCTGATTCGTGTTGCACAGAATCCGGAAGCACCGATGATGCAGGATGATCCGCTG
GCAGCCGGTCTGCGTGATCTGCGTATGCGTGTTGATCGTTTTGGCACCGCAGGCCAGACCGCAC
GTTGGGTTGATGCACTGCGTGAATATTTCTTTAGCGTTGTTTGGGAAGCAGCACATCGTCGTGCA
GGCACCGTTCCGGATCTGAATGATTATACCCTGATGCGTCTGTATGATGGTGCAACCAGCGTTGT
GCTGCCGATGCTGGAAATGGGTCATGGTTATGAACTGCAGCCGTATGAACGTGATCGTACCGCA
GTTCGTGCAGTTGCAGAAATGGCAAGCTTTATTATCACCTGGGATAACGACATCTTCAGCTATCAC
AAAGAACGTCGTGGTAGCGGTTATTATCTGAATGCCCTGCGTGTTCTGGAACAAGAACGTGGTCT
GACACCGGCACAGGCACTGGATGCAGCAATTAGCCAGCGTGATCGTGTTATGTGTCTGTTTACCA
CCGTTAGCGAACAGCTGGCAGAACAGGGTAGCCCTCAGCTGCGTCAGTATCTGCATAGCCTGCG
TTGTTTTATTCGTGGTGCCCAGGATTGGGGTATTAGCAGCGTTCGTTATACCACACCTGATGATCC
GGCAAATATGCCGAGCGTTTTTACCGATGTTCCGACCGATGATAGCACCGAACCGCTGGATATTC
CGGCAGTTAGTTGGTGGTGGGATCTGCTGGCCGAAGATGCACGTAGCGTTCGTCGTCAGGTTCC
GGCACAGCGTAGCGCATAAGAATTCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *rcas_0662* from *R. castenholzii* DSM 13941:

AAAAAGCTAGCAACAAAGAGGAGAGTGCAAAATCGATGGATCAAGACTACCGCGCTCGACTCGTG
TATCCTTTCTCAGGTGCGATCAGCCCGCACGCAGATATCGTGGATCAAGCTACGTTGGCATGGGC
GGCGATGTTCGGACTGCTGACGGACAGTCTGCGCCATAAGTCACGTCGCTTGCAATACGGACTG
TTAGCTGCTCGCGCATACCCTCGAGCGGACCGTGAGATGCTGCAAATCGCAGCAGATTGGATCG
CTTGGTTATTCTTTATGGATGACCAGTGCGACGAAACTGGAATCGGGCGTGATTTGCAGCGCATG
ATCGCGTTACATGAACGATTCCTCGCTATCCACGGCGGAGCTACGCCGGAGGCGCACGATTGCG
CGTTAACGTACGCGTTGGCAGATCTCCGTCGTCGATTAGCTTTACGCGCTCCGGATAATTGGTTA
CGCCGATTTAGCGAGCACGTGCGTTTGTATTTCACTGCAAACCGCTGGGAAACGGTGAATCGTCA
ACGCGGGGCTACGCCGAACGTGGCGACCTACTGCGCAGCGCGCTTATTCAGCGGAGCAGTGTA
CGCTTGCTTTGATTTAATCGAATTAGCTGAGCAAATTGAATTACCGTTTTACGCACGTCATCATAGC
ATCGTGCAACAGTTAGAGCAAGCTGCGAATAACATTATCTGCTGGTGCAACGATGTGTTAAGCTA

CCCTAAAGAAATGCAACATGGAGATCGTCATAACTTAGTGTTAGTGATCCAAGGAGAGCACCAGT
GCAGTCTGCCTGAAGCAATCGACCGCGCGTTGGACCTCCACGCGCGTGAAGTGGCTACGTTCGT
GCGTAAGCGCACGTGTGTGCCTTACTTTGATGCTGCGGTGAACACCGCGCTCGAAAAGTACGTG
ACGGGTCTGCAATTTTGGATTTGCGCAAACCGCGATTGGAGCTTAACGGCTACGCGCTATGCGCC
GACGCATAAGAGCCAAGAAATGGTGATGGCTGTGGCGCAGCAATGAGAATTCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *roseRS_3509* from *R.* sp. RS-1:

AAAAAGCTAGCAACAAAGAGGAGAGTGCAAAATCGATGGATCGTGTGGATGGAGCCCAAATTGTG
TATCCGTTCACGGGGGCGATCAGCCCTTACGCTGGAGATGTGGATCAAGCCACCCTCATTTGGG
CCGAGGCGAGTGGTCTGCTCACGGATGGGCTGCGCCAAAAGTCACAACGGCTTCAATACGGTGT
ATTAGCTGCGCGCGCATACCCTCAAGCCGATCGGGATACGTTACAGATCGCTGCAGATTGGATC
GCATGGTTATTCTTTATGGACGATCAATGTGATGAGGCTGGAATCGGGCGCGATCCGCAACGCTT
AGCTGCGTTACATGAACGATTTCTTGCCGTGCTTGAAGGTGACCCTCCAGATTCAGGAGACTGGA
ACTTAACGCGCGCTCTTGCAGATATCCGCCGGCGTTTAGCAGCGCGCGCCACGGATGATTGGCT
GCGCCGATTTGGTGAGCATGTGCGGCTTTATTTCACGGCGAACCGCTGGGAAGCGGCCAATCGA
TGCCAAAGCATTGTACCTAATGTGGCGACGTACTGTGCTGCGCGCTTATTCAGTGGAGCCGTGTA
CGCATGCTTCGACTTAATCGAATTAGCCGCCGGGATTGATCTCCCTTTCTACGCCCGGTACCACG
CTGCAGTACAGCAGCTTGAGCGTACGGCCAATAACATTATCTGCTGGTGCAACGATATGCTGAGC
TATCCTAAGGAGATGCAACATGGGGATGTGCATAACTTAGTATTAGCTATTCGTCAAGAATACCAG
TGTAGCCTGTCAGAGGCGTTACAACAAGCGCTCTTACTTCATGACCGTGAGACGGACACGTTCAT
GAAGACCCAAAAACAGTTACCGCGCTTTAACCCTGCAGTAGATATGGCGCTCGAGCGCTACATCG
ATGGGCTGCAATATTGGATCTGCGCAAATCGGGACTGGAGTCTGACCGCCATGCGCTACGCTCT
GTCAGGTACGGAACCTGTACTCCGTACGCGTTTCGCCCACAGCTAAGAATTCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prplsWT* promoter:

AAAAAGGTCTCAACAGCTGCAGAACAATCGTTAAAGCGGACGTTTTTGCGCCGCCCGGATTTGCT
TGAAAACTACCCGCTGACAGAAAAGCAAAAACGATGGATCGAAGAGTGGAAAAAAGAAAAACAGT
AGCTATTGCGCATGATACAAGTTTATGCTACTATATTCCTTGTGCAACTTAAGCGATTTGCTTAAGC
GAGGAAAACGATGTTCCGCTGCAATGATGAAAAAGCATTGTCTAGATAAGGAGTGATTCGAATGA
GAGACCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prpls1* promoter:

AAAAAGGTCTCAACAGCTACAGAGCAATCGCTAAAGCGGACGCCTTCGCGCCGCCCGGATTTGC
CTGAGGACTACCCGCTGGCAGAAAAGCAGAAACGACGGATCGAAGAGTGGAAAAAAGAGGAACA
GTAGCTATTGCGCATGATGCGAGTTTATGCTACTATATTCCTTGTGCAACCTAGACGACTTGCTTA
AGCGAGGAAAACGGTGCTCCGCTGCAATGATGAAAAAGCATTGTCTAGATAAGGAGTGATTCGAA
TGAGAGACCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prpls3* promoter:

AAAAAGGTCTCAACAGCTGCAGAACAATCGTTAAAGCGGACGTTTTTGCGCCGCCCGGATTTGCT
TGAAAACTACCCGCTGACAGAAAAGCAAAAACGATGGATCGAAGAGTGGAAAAGAGAAAAACAGT

AGCTATTGCGCATGATACAAGTTTATGCTACTATATTCCTTGTGCAACTTAAGCGATTTGCTTAAGC
GAGGAAAACGATGTTCCGCTGCAATGATGAAAAAGCATTGTCTAGATAAGGAGTGATTCGAATGA
GAGACCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prpls5* promoter:

AAAAAGGTCTCAACAGCTGCAGGACAACCGTTAAAGCGGACGTTTTCGCGCCGCCCGGGTTTGC
TTGAAGACTACCCGCTGACAGAGAAGCAAAGGCGATGGATCGAAGAGTGGAAGAAAGAGAAACA
GTAGCTATTGCGCATGGTACAAGTTTATGCTACTATATTCCTTGCGCAACTTAAGCGATTTGCCTA
AGCGAGGAAAGCGGTGCTCCGCTGCAATGATGAAAAAGCATTGTCTAGATAAGGAGTGATTCGAA
TGAGAGACCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prpls12* promoter:

AAAAAGGTCTCAACAGCTGCAGGACAACCGCTAGGGCGGACACTCCTACGCCGCCCGGACTTGC
TTGGAGACTGCCCGCTGGCCGGAGAGCAGAAGCGACGGCTCGAAGGGTGGAGAAAGGAAAAAC
AGCAGCGATTGCGCATGATACAAGTTTATGCTACTATATCCCTTGTGCAACTTAAGCGGCTTGCTT
AAGTGAGGAAGACGGTGCCCCGCTGCAATGATGAAAGAGCATTGTCTAGATAAGGAGTGATTCGA
ATGAGAGACCAAAAA

The codon optimised DNA sequence for *P. thermoglucosidasius* of *prpls18* promoter:

AAAAAGGTCTCAACAGCTGCAGAACAATCGTTAGAGCGGGTGTTTTCGCGCCGCCCGGACTTGCT
CGAAAGCTACCCGCTGACAGAGAAGCAGAAACGACGGACCGAGGAGTGGAAAAAGGAAAAACAG
TAGCTACTGCGCATGATACAAATCTATGCTACTGTGTTCCTTGTGCAACTTAAGCGGTTTGCTTAA
GCGGGGAAAGCGATGTTCCGCTGCAATGATGGAAAAGCATTGTCTAGATAAGGAGTGATTCGAAT
GAGAGACCAAAAA

Appendix Figure 1: Codon optimised and codon harmonised DNA sequences of TPSs used in this study. All sequences that were codon optimised for *E. coli* contained the NheI (GCTAGC) restriction digestion site at the 5' end and EcoRI (GAATTC) restriction digestion site at the 3' end of the gene. Those genes designed and provided by Dr Matthew Q Styles (University of Bath, UK) also contained a SacI (GAGCTC) restriction digestion site at the 5' end as well as the codon harmonized Rxyl_0493 sequence. Restriction digestion sites are shown in red. Only *roseRS_3509* and *rcas_0662* was codon optimised for *P. thermoglucosidasius* with rare codons removed which was also designed by MQS. The *P. thermoglucosidasius* codon optimised DNA sequences of the constitutive *prpls* promoters were designed by MQS with BsaI restriction digestion sites at both ends of the sequence shown in red.

8.2 Primers created and used in this thesis

Appendix Table 1: Oligonucleotide primers used in this work. All oligonucleotides were designed using Oligoanalyzer (IDT) and synthesised by Eurofins Genomics (Germany).

Name	Sequence (5'-3')	Description
SSCG_For	AAAAAGCTAGCATGAGCCTGAATC	Forward primer for *sscg_03688* truncation to *sscg_R1-R9*
SSCG_Rev1	TATGAATTCTTAGCTCTGCTGCAGACC	Reverse primer for truncation of *sscg_03688* by 7 amino acids to form *sscg_R1*
SSCG_Rev2	TATGAATTCTTAGCTTGCGGTAATGCCA	Reverse primer for truncation of *sscg_03688* by 15 amino acids to form *sscg_R2*
SSCG_Rev3	TATGAATTCTTAGGTGCGCTGATGATGC	Reverse primer for truncation of *sscg_03688* by 29 amino acids to form *sscg_R3*
SSCG_Rev4	TATGAATTCTTAGGTAGGCGGACGATCA	Reverse primer for truncation of *sscg_03688* by 41 amino acids to form *sscg_R4*
SSCG_Rev5	TATGAATTCTTAACGTTTACGAACCGGTGT	Reverse primer for truncation of *sscg_03688* by 49 amino acids to form *sscg_R5*
SSCG_Rev6	TATGAATTCTTAACCATCAGAACCCGGTT	Reverse primer for truncation of *sscg_03688* by 57 amino acids to form *sscg_R6*
SSCG_Rev7	TATGAATTCTTACCACCACCAACCAATAACATC	Reverse primer for truncation of *sscg_03688* by 70 amino acids to form *sscg_R7*
SSCG_Rev8	TATGAATTCTTACGGACCCTGACCACTC	Reverse primer for truncation of *sscg_03688* by 82 amino acids to form *sscg_R8*
SSCG_Rev9	TATGAATTCTTAAACTTCATGCGGACTCAGG	Reverse primer for truncation of *sscg_03688* by 90 amino acids to form *sscg_R9*
T7 Forward	TAATACGACTCACTATAGGG	Forward primer for amplifying or sequencing the multiple cloning site of pET28a
T7 Reverse	CTAGTTATTGCTCAGCGGT	Reverse primer for amplifying or sequencing the multiple cloning site of pET28a
SSCG_R6_SDM_F1	ATGCCATACCGCGAAAGG	Forward primer for introducing mutations into *sscg_R6* from pET28a
SSCG_R6_SDM_F2	GGTGATGTCGGCGATATAGG	Forward primer for introducing mutations into *sscg_R6* from pET28a
SSCG_R6_SDM_R1	AAGCTTGTCGACGGAGC	Reverse primer for introducing mutations into *sscg_R6* from pET28a
SSCG_R6_SDM_R2	CCAGTTTGGAACAAGAGTCCAC	Reverse primer for introducing mutations into *sscg_R6* from pET28a
SSCG_R6_SDM_R3	CCCGATTTAGAGCTTGACGG	Reverse primer for introducing mutations into *sscg_R6* from pET28a
SSCG_R6_G24I_IR	TTAACAATGCTTGCTGCCGGATGCGGCAGATCAT	Internal reverse primer for introducing the G24I mutation into *sscg_R6* from pET28a
SSCG_R6_G24I_IF	GCAGCAAGCATTGTTAATGATCGTACCCTGGATTGG	Internal forward primer for introducing the G24I mutation into *sscg_R6* from pET28a

SSCG_R6_P39I_IR	TGCGGTTATAATACCCTGACCGCTTGCCCAATCCAGG	Internal reverse primer for introducing the P39I mutation into *sscg_R6* from pET28a
SSCG_R6_P39I_IF	GTCAGGGTATTATAACCGCAGATCGTGATGCAGGTCGTC	Internal forward primer for introducing the P39I mutation into *sscg_R6* from pET28a
SSCG_R6_S100Q_IR	AGACGTGCCAGAAATTGTGCCCATTCGGTAATACCGCTACCATCTGCA	Internal reverse primer for introducing the S100Q mutation into *sscg_R6* from pET28a
SSCG_R6_S100Q_IF	CGAATGGGCACAATTTCTGGCACGTCTGCATCGTGTTGTTGAAACCG	Internal forward primer for introducing the S100Q mutation into *sscg_R6* from pET28a
SSCG_R6_A115P_IR	TACCAGGCAGCAGTGGGCTTTCACCGGTTTCAACAACACG	Internal reverse primer for introducing the A115P mutation into *sscg_R6* from pET28a
SSCG_R6_A115P_IF	AAAGCCCACTGCTGCCTGGTAATCCGTATGGTCTGGCAC	Internal forward primer for introducing the A115P mutation into *sscg_R6* from pET28a
SSCG_R6_G209A_IR	AGTGCACGAACAGCCGGTGTTTCACGCAGTTCACGAGG	Internal reverse primer for introducing the G209A mutation into *sscg_R6* from pET28a
SSCG_R6_G209A_IF	CACCGGCTGTTCGTGCACTGGTTGAAATGACAGCACTGC	Internal forward primer for introducing the G209A mutation into *sscg_R6* from pET28a
SSCG_R6_D224C_IR	TCAGAATATCATTACAAACGCTAACCAGCAGTGCTGTCATTTCAACCAG	Internal reverse primer for introducing the D224C mutation into *sscg_R6* from pET28a
SSCG_R6_D224C_IF	CTGGTTAGCGTTTGTAATGATATTCTGAGCCATCACAAAGAAAGCACC	Internal forward primer for introducing the D224C mutation into *sscg_R6* from pET28a
SSCG_R6_T293L_IR	GTGCCAGGCCTGTGAGAAAACGATACAGTTCCGGCACG	Internal reverse primer for introducing the T293L mutation into *sscg_R6* from pET28a
SSCG_R6_T293L_IF	TTTTCTCACAGGCCTGGCACGTTGGATTCGTGCAAATCTGG	Internal forward primer for introducing the T293L mutation into *sscg_R6* from pET28a
TS03_L54A_IR	CGCGCGCAGCTGCTACACCGTATTGAAGCCGTTGTGACTTTTGGC	Internal reverse primer for introducing the L54A mutation into *roseRS_3509* from pET28a

TS03_L54A_IF	CAATACGGTGTAGCAGCTGCGCGCGCATACCCTCAAGCCGATCG	Internal forward primer for introducing the L54A mutation into *roseRS_3509* from pET28a
TS03_Y59A_IR	ATCGGCTTGAGGGGCTGCGCGCGCAGCTAATACACCGTATTGAAGCCG	Internal reverse primer for introducing the Y59A mutation into *roseRS_3509* from pET28a
TS03_Y59A_IF	TGCGCGCGCAGCCCCTCAAGCCGATCGGGATACGTTACAGATCGC	Internal forward primer for introducing the Y59A mutation into *roseRS_3509* from pET28a
TS03_I74A_IR	ATAAAGAATAACCATGCGGCCCAATCTGCAGCGATCTGTAACGTATCCCGA	Internal reverse primer for introducing the I74A mutation into *roseRS_3509* from pET28a
TS03_I74A_IF	TGCAGATTGGGCCGCATGGTTATTCTTTATGGACGATCAATGTGATGAGGC	Internal forward primer for introducing the I74A mutation into *roseRS_3509* from pET28a
TS03_C184A_IR	TCGAAGGCTGCGTACACGGCTCCACTGAATAAGCGC	Internal reverse primer for introducing the C184A mutation into *roseRS_3509* from pET28a
TS03_C184A_IF	GTGTACGCAGCCTTCGACTTAATCGAATTAGCCGCCGG	Internal forward primer for introducing the C184A mutation into *roseRS_3509* from pET28a
TS03_W296L_IR	TGCGCAGATCAAATATTGCAGCCCATCGATGTAGCGCTC	Internal reverse primer for introducing the W296L mutation into *roseRS_3509* from pET28a
TS03_W296L_IF	CTGCAATATTTGATCTGCGCAAATCGGGACTGGAGTCTGAC	Internal forward primer for introducing the W296L mutation into *roseRS_3509* from pET28a
TS03_W296A_IR	TGCGCAGATCGCATATTGCAGCCCATCGATGTAGCGCTC	Internal reverse primer for introducing the W296A mutation into *roseRS_3509* from pET28a
TS03_W296A_IF	CTGCAATATGCGATCTGCGCAAATCGGGACTGGAGTCTGAC	Internal forward primer for introducing the W296A mutation into *roseRS_3509* from pET28a
TS03_L77A_IR	AAGAATGCCCATGCGATCCAATCTGCAGCG	Internal reverse primer for introducing the L77A mutation into *roseRS_3509* from pET28a
TS03_L77A_IF	ATGGGCATTCTTTATGGACGATCAATGTGATGAGGC	Internal forward primer for introducing the L77A mutation into *roseRS_3509* from pET28a

TS03_F78A_IR	CATTGATCGTCCATAAAGGCTAACCATGCGATCCAATCTGCAGCG	Internal reverse primer for introducing the F78A mutation into *roseRS_3509* from pET28a
TS03_F78A_IF	ATGGTTAGCCTTTATGGACGATCAATGTGATGAGGCTGGAATCGGG	Internal forward primer for introducing the F78A mutation into *roseRS_3509* from pET28a
TS03_F78V_IR	CATTGATCGTCCATAAAGACTAACCATGCGATCCAATCTGCAGCG	Internal reverse primer for introducing the F78V mutation into *roseRS_3509* from pET28a
TS03_F78V_IF	ATGGTTAGTCTTTATGGACGATCAATGTGATGAGGCTGGAATCGGG	Internal forward primer for introducing the F78V mutation into *roseRS_3509* from pET28a
TS03_N300A_IR	CAGTCCCGAGCTGCGCAGATCCAATATTGCAGCCCATCGA	Internal reverse primer for introducing the N300A mutation into *roseRS_3509* from pET28a
TS03_N300A_IF	TCTGCGCAGCTCGGGACTGGAGTCTGACCGCCATGC	Internal forward primer for introducing the N300A mutation into *roseRS_3509* from pET28a
TS03_W303A_IR	CATGGCGGTCAGACTCGCGTCCCGATTTGCGCAGATCCAATATTGCAGCCC	Internal reverse primer for introducing the W303A mutation into *roseRS_3509* from pET28a
TS03_W303A_IF	GCAAATCGGGACGCGAGTCTGACCGCCATGCGCTACGCTCTGTCAGG	Internal forward primer for introducing the W303A mutation into *roseRS_3509* from pET28a
TS03_Y310A_IR	CGTACCTGACAGAGCGGCGCGCATGGCGGTCAGACTCC	Internal reverse primer for introducing the Y310A mutation into *roseRS_3509* from pET28a
TS03_Y310A_IF	CGCGCCGCTCTGTCAGGTACGGAACCTGTACTCCGTACGCG	Internal forward primer for introducing the Y310A mutation into *roseRS_3509* from pET28a
SSCG_G46H_IR	CACGCAGACGATGTGCATCACGATCTGCGGTCGGAATACCCTG	Internal reverse primer for introducing the G46H mutation into *sscg_R6* from pET28a
SSCG_G46H_IF	AGATCGTGATGCACATCGTCTGCGTGCAATGGCACCGGGTCTG	Internal forward primer for introducing the G46H mutation into *sscg_R6* from pET28a
SSCG_H106L_IR	TTTCAACAACACGAAGCAGACGTGCCAGAAAGCTTGCCCAT	Internal reverse primer for introducing the H106L mutation into *sscg_R6* from pET28a

SSCG_H106L_IF	ACGTCTGCTTCGTGTTGTTGAAACCGGTGAAAGCGCACT	Internal forward primer for introducing the H106L mutation into *sscg_R6* from pET28a
SSCG_T207H_IR	CGAACACCCGGGTGTTCACGCAGTTCACGAGGCAGAAC	Internal reverse primer for introducing the T207H mutation into *sscg_R6* from pET28a
SSCG_T207H_IF	GTGAACACCCGGGTGTTCGTGCACTGGTTGAAATGACAGC	Internal forward primer for introducing the T207H mutation into *sscg_R6* from pET28a
SSCG_V274M_IR	GTTCTGCAACACGCATAAACTGACGCATAATTTCATTGCGCAGGG	Internal reverse primer for introducing the V274M mutation into *sscg_R6* from pET28a
SSCG_V274M_IF	GTCAGTTTATGCGTGTTGCAGAACGTGTTCGTACACCGGCAG	Internal forward primer for introducing the V274M mutation into *sscg_R6* from pET28a
SSCG_T311L_IR	GGACCGGTATAACGGAGGGTGGTCAGTGAAAAATCCAGATTTGCACGAATCC	Internal reverse primer for introducing the T311L mutation into *sscg_R6* from pET28a
SSCG_T311L_IF	CTGACCACCCTCCGTTATACCGGTCCGGTTACCGAACGTGCAGC	Internal forward primer for introducing the T311L mutation into *sscg_R6* from pET28a
SDS_G45K_IR	GGTAACCAGTTTCTTACGCAGTTCTTCGCTACCAATACG	Internal reverse primer for introducing the G45K mutation into *sds* from pET28a
SDS_G45K_IF	GTAAGAAACTGGTTACCCAGGATATTGGCACCTTTAGCG	Internal forward primer for introducing the G45K mutation into *sds* from pET28a
SDS_I106L_IR	CAACACGAAGCAGACGATGCAGCAGACCTGCCAGATCACCC	Internal reverse primer for introducing the I106L mutation into *sds* from pET28a
SDS_I106L_IF	GCTGCATCGTCTGCTTCGTGTTGCACAGAATCCGGAAGCACC	Internal forward primer for introducing the I106L mutation into *sds* from pET28a
SDS_R206H_IR	GAACTGCGGTATGATCACGTTCATACGGCTGCAGTTCATAACC	Internal reverse primer for introducing the R206H mutation into *sds* from pET28a
SDS_R206H_IF	AACGTGATCATACCGCAGTTCGTGCAGTTGCAGAAATGGC	Internal forward primer for introducing the R206H mutation into *sds* from pET28a

SDS_T273K_IR	CGCTAACGGTCTTAAACAGACACATAACACGATCACGCTGGC	Internal reverse primer for introducing the T273K mutation into *sds* from pET28a
SDS_T273K_IF	GTGTCTGTTTAAGACCGTTAGCGAACAGCTGGCAGAACAGG	Internal forward primer for introducing the T273K mutation into *sds* from pET28a
SDS_V309P_IR	TATAACGAGGGCTGCTAATACCCCAATCCTGGGCACCA	Internal reverse primer for introducing the V309P mutation into *sds* from pET28a
SDS_V309P_IF	GGTATTAGCAGCCCTCGTTATACCACACCTGATGATCCGG	Internal forward primer for introducing the V309P mutation into *sds* from pET28a
M13 -20 For	GTAAAACGACGGCCAG	Forward primer for amplifying or sequencing from the pMA plasmids
M13 -R -29	CAGGAAACAGCTATGAC	Reverse primer for amplifying or sequencing from the pMA plasmids
pG2AC_oriT_SDM_MCS_For	GCCCCGTTAGTTGAAGAAGGT	Forward primer for amplifying or sequencing from pG2AC oriT SDM
pG2AC/pG2K_MCS_Rev	GCGGATAACAATTTCACACAGG	Reverse primer for amplifying or sequencing from pG2AC oriT SDM or pG2K oriT SDM bgl sf-gfp GGready
pG2K_oriT_SDM_bgl_sf-gfp_GGready_MCS_For	CCTCTACGGACCGAGTATTC	Forward primer for amplifying or sequencing from pG2K oriT SDM bgl sf-gfp GGready
GeoOptW296A_IR	GCGCAAATCGCATATTGTAATCCATCAATATAGCGTTCTAACG	Internal reverse primer for introducing the W296A mutation into *Parageobacillus* optimised *roseRS_3509* in pG2K oriT SDM bgl sf-gfp GGready
GeoOptW296A_IF	AATATGCGATTTGCGCGAATCGCGATTGGAGCTTAACG	Internal forward primer for introducing the W296A mutation into *Parageobacillus* optimised *roseRS_3509* in pG2K oriT SDM bgl sf-gfp GGready
JKG1_TAA_Rev	AAAAAGAATTCTTACAGACACATGGTTTCCGG	Reverse primer to add a stop codon to JKG1 in pET28a
MEV_F1	AACAGACGAACGTGAACAG	Forward sequencing primer of the *hmgr* gene in P2 Mev9
MEV_F2	GAGCAAATATCTGTATGATGCG	Forward sequencing primer of the *aact* gene in P2 Mev9
MEV_F3	AGCATTAACGCGATTAAACGC	Forward sequencing primer of the *hmgs* gene in P2 Mev9
MEV_F4	GAGCTATGCGGAATATGCG	Forward sequencing primer of the *hmgs* gene in P2 Mev9

MEV_F5	GTACGTGAATCATGGATTACTG	Forward sequencing primer of the *mk* gene in P2 Mev9
MEV_F6	CATAGCAATAGCATGCACG	Forward sequencing primer of the *dmd* gene in P2 Mev9
MEV_F7	GGAGAATTATATGGCGATAAAACTG	Forward sequencing primer of the *pmk* gene in P2 Mev9
MEV_F8	CATGGAAACGGCGAAATTATTATATAG	Forward sequencing primer of the *idi* gene in P2 Mev9
MEV_F9	CGGAAGGCATGGTGG	Forward sequencing primer of the *fpps* gene in P2 Mev9
MEV_R0	GCGTTAAACTGGCTCAG	Reverse sequencing primer of the *hmgr* gene in P2 Mev9
MEV_R1	GCCAGCTGATAAATCGC	Reverse sequencing primer of the *aact* gene in P2 Mev9
MEV_R2	CAATATACGGGCTCACCAG	Forward sequencing primer of the *hmgs* gene in P2 Mev9
MEV_R3	GACCTCCATACGTTTCCG	Forward sequencing primer of the *mk* gene in P2 Mev9
MEV_R4	CCTTATTCCACACCACAAATC	Forward sequencing primer of the *dmd* gene in P2 Mev9
MEV_R5	GCGCCAGTTTATGAATCTC	Forward sequencing primer of the *pmk* gene in P2 Mev9
MEV_R6	GCGCACAATCGTAAAGC	Forward sequencing primer of the *idi* gene in P2 Mev9
MEV_R7	GCGTTCAATATAGCGGC	Forward sequencing primer of the *fpps* gene in P2 Mev9
Gib_p*rplsWT/3*_For	CGTACTAGTCCTAGGCTAGCGACAGCTGCAGAACAATCGTTAAAGCGG	Gibson Assembly forward primer for *prplsWT/3* amplification
Gib_*prplsWT/3*_Rev	TGCGCTCCATCCACGCGATCCATTCGAATCACTCCTTATCTAGACAATGC	Gibson Assembly reverse primer for *prplsWT/3* amplification
Gib_GeoOptW296A_For	CTAGATAAGGAGTGATTCGAATGGATCGCGTGGATGGAG	Gibson Assembly forward primer for *geoopt_W296A* amplification
Gib_GeoOptW296A_Rev	GCGCGAAATTGAGCTCAGTGAAGCTTCTAGTCTAGAACACTCCTGC	Gibson Assembly reverse primer for *geoopt_W296A* amplification
Gib_pG2ACoriTSDM_For	GCAGGAGTGTTCTAGACTAGAAGCTTCACTGAGCTCAATTTCG	Gibson Assembly forward primer for the pG2AC oriT SDM backbone amplification
Gib_pG2ACoriTSDM_Rev	CTTTAACGATTGTTCTGCAGCTGTCGCTAGCCTAGGACTAGTACGC	Gibson Assembly reverse primer for the pG2AC oriT SDM backbone amplification

GG_*prplsWT/3*_For	AAAAAGGTCTCAACAGCTGCAGAACAATCGTTAAAGCGGA	Forward primer to introduce Golden Gate sites for the *prplsWT/3* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls5*_For	AAAAAGGTCTCAACAGCTGCAGGACAACCGTTAAAGC	Forward primer to introduce Golden Gate sites for the *prpls5* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_p*rpls12*_For	AAAAAGGTCTCAACAGCTGCAGGACAACCGCTAGG	Forward primer to introduce Golden Gate sites for the *prpls12* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls18*_For	AAAAAGGTCTCAACAGCTGCAGAACAATCGTTAGAGCG	Forward primer to introduce Golden Gate sites for the *prpls18* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls1*v1_For	AAAAAGGTCTCAACAGCTACAGAGCAATCGCTAAAGC	Forward primer to introduce Golden Gate sites for the *prpls1* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls1*v1_Rev	AAAAAGGTCTCACATTCGAATCACTCCTTATCTAGAC	Reverse primer to introduce Golden Gate sites for the *prpls1* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls1*v2_For	AAAAAGGTCTCAACAGctacagAGCAATCGCTAAAGCGG	Forward primer to introduce Golden Gate sites for the *prpls1* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls1*v2_Rev	AAAAAGGTCTCACATTCGAATCACTCCTTATCTAGACAATGC	Reverse primer to introduce Golden Gate sites for the *prpls1* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_*prpls*_Rev	AAAAAGGTCTCACATTCGAATCACTCCTTATCTAGACAATGC	Reverse primer to introduce Golden Gate sites for the *prpls* promoter insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_GeoOpt_W296A_For	AAAAAGGTCTCAAATGGATCGCGTGGATGG	Forward primer to introduce Golden Gate sites for the *geoopt_W296A* insertion into pG2K oriT SDM bgl sf-gfp GGready
GG_GeoOpt_W296A_Rev	AAAAAGGTCTCACTAGTTAGCTATGCGCAAAGCG	Reverse primer to introduce Golden Gate sites for the *geoopt_W296A* insertion into pG2K oriT SDM bgl sf-gfp GGready
AmpR_For	ATTAAGGCCGGCCCTGACAGTTACCAATGCTTAATCAGTG	Forward primer for amplifying *ampR* from pG2AC oriT SDM
AmpR_Rev	TATAAAAATAGGCGTATCACGAGGCCGCGGCCGCAAA	Reverse primer for amplifying *ampR* from pG2AC oriT SDM

QC_*prpls18*_W296A_For	ATATTGATGGATTACAATATGCGATTTGCGCGAATCGCGATTG	Forward primer for Quikchange of pG2AC oriT SDM *prpls18 geoopt_W296A* construct
QC_*prpls18*_W296A_Rev	CGCATATTGTAATCCATCAATATAGCGTTCTAACGCCATATCC	Reverse primer for Quikchange of pG2AC oriT SDM *prpls18 geoopt_W296A* construct
Gib_*prpls18*_For	CGTACTAGTCCTAGGCTAGCGACAGCTGCAGAACAATCGTTAGAGCGG	Forward Gibson Assembly primer for the *prpls18* fragment in the pG2AC oriT SDM vector
Gib_KO_repB_For	GGCCATTGGTAACTGTCAGACCGAAGTCGAGATCAGGGAATGAG	Forward Gibson Assembly primer for repB fragment in the Knockout vector
Gib_KO_repB_Rev	TAGTTGAAGAATAAAAGACCGATCCTCTAGAGTCGACCTGC	Reverse Gibson Assembly primer for repB fragment in the Knockout vector
Gib_KO_ChlR_For	GCAGGTCGACTCTAGAGGATCGGTCTTTTATTCTTCAACTAAAGCACCC	Forward Gibson Assembly primer for c*hlR* fragment in the Knockout vector
Gib_KO_ChlR_Rev	CGCCTTTCACCAGTTTTTCCAGCTAGCCTAGGACTAGTACGC	Reverse Gibson Assembly primer for c*hlR* fragment in the Knockout vector
Gib_KO_tHMGR_For	CGTACTAGTCCTAGGCTAGCTGGAAAAACTGGTGAAAGGCG	Forward Gibson Assembly primer for *thmgr* fragment in the Knockout vector
Gib_KO_tHMGR_Rev	ATATAAGTTGTTACTAGTGCAGTTTCGCATGCGCTTTGC	Reverse Gibson Assembly primer for *thmgr* fragment in the Knockout vector
Gib_KO_p15A_For	GCAAAGCGCATGCGAAACTGCACTAGTAACAACTTATAT	Forward Gibson Assembly primer for p15A fragment in the Knockout vector
Gib_KO_p15A_Rev	GCCACCTGACGTCTAAGAAACGGAAATGGCTTACGAACGGG	Reverse Gibson Assembly primer for p15A fragment in the Knockout vector
Gib_KO_AmpR_For	CCCGTTCGTAAGCCATTTCCGTTTCTTAGACGTCAGGTGGC	Forward Gibson Assembly primer for a*mpR* fragment in the Knockout vector
Gib_KO_AmpR_Rev	CTCATTCCCTGATCTCGACTTCGGTCTGACAGTTACCAATGGCC	Reverse Gibson Assembly primer for a*mpR* fragment in the Knockout vector
Gib_KI_repB_For	CCCGTTCGTAAGCCATTTCCCGAAGTCGAGATCAGGGAATGAG	Forward Gibson Assembly primer for repB fragment in the Knock-in vector
Gib_KI_repB_Rev	GTTCATCCATACCATGCGTGGATCCTCTAGAGTCGACCTGC	Reverse Gibson Assembly primer for repB fragment in the Knock-in vector
Gib_KI_KanR/GeoOptW296A_For	CAGGTCGACTCTAGAGGATCCACGCATGGTATGGATGAACTG	Forward Gibson Assembly primer for *kanR*/*geooptW296A* fragment in the Knock-in vector

Gib_KI_KanR/GeoOptW296A _Rev	TTATTCATCATGATCGACAGGTGTCTAGTCTAGTTAGCTATGCGC	Reverse Gibson Assembly primer for *kanR*/*geooptW296A* fragment in the Knock-in vector
Gib_KI_tHMGR_For	TAGCTAACTAGACTAGACACCTGTCGATCATGATGAATAAGAAGTTGGG	Forward Gibson Assembly primer for *thmgr* fragment in the Knock-in vector
Gib_KI_tHMGR_Rev	ATATAAGTTGTTACTAGTGCAGTTTCGCATGCGCTTTGC	Reverse Gibson Assembly primer for *thmgr* fragment in the Knock-in vector
Gib_KI_p15A_For	GGCAAAGCGCATGCGAAACTGCACTAGTAACAACTTATATCGTATGGG	Forward Gibson Assembly primer for p15A fragment in the Knock-in vector
Gib_KI_p15A_Rev	ATTCCCTGATCTCGACTTCGGGAAATGGCTTACGAACGGG	Reverse Gibson Assembly primer for p15A fragment in the Knock-in vector

8.3 Plasmids used and/or created in this thesis

Appendix Table 2: Plasmids used in this work. The genes ordered from GeneArt® (Thermo Fisher, UK) arrived in the pMA vector. The rest of the plasmids were made in this study. *According to MQS (University of Bath, UK).

Plasmid Name	Description	Source
pCR:*tcur_3107*	Holding vector for *tcur_3107* made by blunt cloning into pCR™-Blunt.	This Study
pCR:*fjsc11*	Holding vector for fjsc*11* made by blunt cloning into pCR™-Blunt.	This Study
pCR:*tchrom*	Holding vector for *tchrom* made by blunt cloning into pCR™-Blunt.	This Study
pMA:*tbis_3257*	The vector containing the *E. coli* optimised *tbis_3257* synthesised by GeneArt.	GeneArt, Thermo Fisher
pCR:*jkg1*	Holding vector for *jkg1* made by blunt cloning into pCR™-Blunt.	This Study
pMA:*rxyl_0493*	The vector containing the *E. coli* optimised *rxyl_0493* synthesised by GeneArt.	GeneArt, Thermo Fisher
pET28a:*tcur_3107*	The *E. coli* expression vector for Tcur_3107 inserted into the MCS.	This Study
pET28a:*fjsc11*	The *E. coli* expression vector for FJSC11 inserted into the MCS.	This Study
pET28a:*tchrom*	The *E. coli* expression vector for Tchrom inserted into the MCS.	This Study
pET28a:*tbis_3257*	The *E. coli* expression vector for Tbis_3257 inserted into the MCS.	This Study

pET28a:*jkg1*	The *E. coli* expression vector for JKG1 inserted into the MCS.	This Study
pET28a:*rxyl*_0493	The *E. coli* expression vector for Rxyl_0493 inserted into the MCS.	This Study
pMA:*sscg_03688*	The vector containing the *E. coli* optimised SSCG_03688 synthesised by GeneArt.	GeneArt, Thermo Fisher
pET28a:*sscg_03688*	The *E. coli* expression vector for SSCG_03688 inserted into the MCS.	This Study
pET28a:*sscg_R1*	The *E. coli* expression vector for SSCG_R1 inserted into the MCS.	This Study
pET28a:*sscg_R2*	The *E. coli* expression vector for SSCG_R2 inserted into the MCS.	This Study
pET28a:*sscg_R3*	The *E. coli* expression vector for SSCG_R3 inserted into the MCS.	This Study
pET28a:*sscg_R4*	The *E. coli* expression vector for SSCG_R4 inserted into the MCS.	This Study
pET28a:*sscg_R5*	The *E. coli* expression vector for SSCG_R5 inserted into the MCS.	This Study
pET28a:*sscg_R6*	The *E. coli* expression vector for SSCG_R6 inserted into the MCS.	This Study
pET28a:*sscg_R7*	The *E. coli* expression vector for SSCG_R7 inserted into the MCS.	This Study
pET28a:*sscg_R8*	The *E. coli* expression vector for SSCG_R8 inserted into the MCS.	This Study
pET28a:*sscg_R9*	The *E. coli* expression vector for SSCG_R9 inserted into the MCS.	This Study
pET28a:*sscg_R6_G24I*	The *E. coli* expression vector for SSCG_R6_G24I inserted into the MCS.	This Study
pET28a:*sscg_R6_P39I*	The *E. coli* expression vector for SSCG_R6_P39I inserted into the MCS.	This Study
pET28a:*sscg_R6_S100Q*	The *E. coli* expression vector for SSCG_R6_S100Q inserted into the MCS.	This Study
pET28a:*sscg_R6_A115P*	The *E. coli* expression vector for SSCG_R6_A115P inserted into the MCS.	This Study
pET28a:*sscg_R6_G209A*	The *E. coli* expression vector for SSCG_R6_G209A inserted into the MCS.	This Study
pET28a:*sscg_R6_D224C*	The *E. coli* expression vector for SSCG_R6_D224C inserted into the MCS.	This Study
pET28a:*sscg_R6_T293L*	The *E. coli* expression vector for SSCG_R6_T293L inserted into the MCS.	This Study
pET28a:*sscg_R6_G46H*	The *E. coli* expression vector for SSCG_R6_G46H inserted into the MCS.	This Study
pET28a:*sscg_R6_H106L*	The *E. coli* expression vector for SSCG_R6_H106L inserted into the MCS.	This Study
pET28a:*sscg_R6_T207H*	The *E. coli* expression vector for SSCG_R6_T207H inserted into the MCS.	This Study
pET28a:*sscg_R6_V274M*	The *E. coli* expression vector for SSCG_R6_V274M inserted into the MCS.	This Study
pET28a:*sscg_R6_T311L*	The *E. coli* expression vector for SSCG_R6_T311L inserted into the MCS.	This Study

pET28a:*sds*	The *E. coli* expression vector for SdS inserted in the MCS.	This Study
pET28a:*sds_G45K*	The *E. coli* expression vector for SdS_G45K inserted into the MCS.	This Study
pET28a:*sds_I106L*	The *E. coli* expression vector for SdS_I106L inserted into the MCS.	This Study
pET28a:*sds_R206H*	The *E. coli* expression vector for SdS_R206H inserted into the MCS.	This Study
pET28a:*sds_T273K*	The *E. coli* expression vector for SdS_T273K inserted into the MCS.	This Study
pET28a:*sds_V309P*	The *E. coli* expression vector for SdS_V309P inserted into the MCS.	This Study
pMA:*roseRS_3509*	The vector containing the *E. coli* optimised RoseRS_3509 synthesised by GeneArt.	GeneArt, Thermo Fisher
pMA:*rcas_0662*	The vector containing the *E. coli* optimised Rcas_0662 synthesised by GeneArt.	GeneArt, Thermo Fisher
pET28a:*roseRS_3509*	The *E. coli* expression vector for RoseRS_3509 inserted into the MCS.	This Study
pET28a:*rcas_0662*	The *E. coli* expression vector for Rcas_0662 inserted into the MCS.	This Study
pET28a:*roseRS_3509_L54A*	The *E. coli* expression vector for RoseRS_3509_L54A inserted into the MCS.	This Study
pET28a:*roseRS_3509_Y59A*	The *E. coli* expression vector for RoseRS_3509_Y59A inserted into the MCS.	This Study
pET28a:*roseRS_3509_L77A*	The *E. coli* expression vector for RoseRS_3509_L77A inserted into the MCS.	This Study
pET28a:*roseRS_3509_F78A*	The *E. coli* expression vector for RoseRS_3509_F78A inserted into the MCS.	This Study
pET28a:*roseRS_3509_F78V*	The *E. coli* expression vector for RoseRS_3509_F78V inserted into the MCS.	This Study
pET28a:*roseRS_3509_C184A*	The *E. coli* expression vector for RoseRS_3509_C184A inserted into the MCS.	This Study
pET28a:*roseRS_3509_W296A*	The *E. coli* expression vector for RoseRS_3509_W296A inserted into the MCS.	This Study
pET28a:*roseRS_3509_W296L*	The *E. coli* expression vector for RoseRS_3509_W296L inserted into the MCS.	This Study
pET28a:*roseRS_3509_N300A*	The *E. coli* expression vector for RoseRS_3509_N300A inserted into the MCS.	This Study
pET28a:*roseRS_3509_W303A*	The *E. coli* expression vector for RoseRS_3509_W303A inserted into the MCS.	This Study
pET28a:*roseRS_3509_Y310A*	The *E. coli* expression vector for RoseRS_3509_Y310A inserted into the MCS.	This Study
pG2K_oriT_SDM_bgl_sf-gfp_GGready:*prpls1-geoOpt_W296A*	The holding vector for *prpls1* g*eoopt_W296A* made by GoldenGate. g*eoopt_W296A*.	This Study
pG2K_oriT_SDM_bgl_sf-gfp_GGready:*prpls5-geoOpt_W296A*	The holding vector for *prpls5* g*eoopt_W296A* made by GoldenGate. g*eoopt_W296A*.	This Study

pG2K_oriT_SDM_bgl_sf-gfp_GGready:*prpls12- geoOpt_W296A*	The holding vector for *prpls12* g*eoopt_W296A* made by GoldenGate. g*eoopt_W296A*.	This Study
pG2K_oriT_SDM_bgl_sf-gfp_GGready:*prpls18- geoOpt_W296A*	The holding vector for *prpls18* g*eoopt_W296A* made by GoldenGate. *geoopt_W296A*.	This Study
pG2AC_oriT_SDM:*prpls1- geoOpt_W296A*	The *Parageobacillus* vector for expression of *geoOpt_W296A* under the *prpls1* promoter.	This Study
pG2AC_oriT_SDM:*prpls5- geoOpt_W296A*	The *Parageobacillus* vector for expression of *geoOpt_W296A* under the *prpls5* promoter.	This Study
pG2AC_oriT_SDM:*prpls12- geoOpt_W296A*	The *Parageobacillus* vector for expression of *geoOpt_W296A* under the *prpls12* promoter.	This Study
pUM3	The *P. thermoglucosidasius* vector containing the upper mevalonate (UM3) pathway genes under the control of the maltose inducible promoter (*pglv*); AACT, HMGS and HMGR.	MQS
pLM5	The *P. thermoglucosidasius* vector containing the lower mevalonate (LM5) pathway genes under the control of the maltose inducible promoter (*pglv*): MK, PMK, DMD, IDI and FPPS.	MQS
pUC57:*prplsWT**	The holding vector for the *prplsWT* constitutive promoter	MQS
pUC57:*prpls1**	The holding vector for the *prpls1* constitutive promoter	MQS
pUC57:*prpls3**	The holding vector for the *prpls3* constitutive promoter	MQS
pUC57:*prpls5**	The holding vector for the *prpls5* constitutive promoter	MQS
pUC57:*prpls12**	The holding vector for the *prpls12* constitutive promoter	MQS
pUC57:*prpls18**	The holding vector for the *prpls18* constitutive promoter	MQS

8.4 HMM similarity and identity tables between multiple sequence alignment sequences

Appendix Figure 2: Percentage similarity and identity of the TPS amino acid sequences used in the multiple sequence alignment for the A) class I TPS HMM and B) class II TPS HMM. C) The sliding colour scale for the % similarity of identity. The amino acids considered as similar to one another were GAVLI, FYW, CM, ST, KRH, DENQ and P using the online Ident and Sim server on the Sequence Manipulation Suite.

8.5 Sequence alignments for class I/II terpene synthase HMMs

8.5.1 Class I terpene synthase sequence alignment

```
2-Methylisoborneol   MKDTNLDNTSTTFLFGPTGLGTSAARFVSEIAKAARTDNLDSGLLQTDLA
4-epi-cubebol        MLD-------------------------------------------------
African-1-ene        MPF-------------------------------------------------
Alpha-amorphene      MSTTHEEIA----LAGPDGI--------------------------------
Cembrene_C           MTDPAVT-----------P---------------------------------
Corvol_Ether_B       MIPR------------------------------------------------
Delta-cadinene       MSTRPVE---------------------------------------------
Hedycaryol           MHIETDYG--------------------------------------------
Intermedeol          MNPRMT----------------------------------------------
Linalool             MQEFE-----------------------------------------------
Odyverdiene_A        MPFV------------------------------------------------
Pimaradiene          MRARHRVA--------------------------------------------
Tau-muurolol         MDRVD-----------------------------------------------
Terpentetriene       MPDAI-----------------------------------------------
_-_-Isohirsut-4-ene  MTTTAEILS-------P--------------------------------
                     *

2-Methylisoborneol   SVGDCDPAYAERDWGDGTASPL-YCPITER-FNEPLADEVDDRLAVWALE
4-epi-cubebol        ------------------------------------------------------
African-1-ene        ---------------------------PARPL-NAGMEQARQRMWAWIDH
Alpha-amorphene      ------PAV---DLRDLIDAQL-YMPFPFE-R-NPHASEAAAGVDHWLST
Cembrene_C           --------------LAFSIPQL-YCPFPTA-I-HPEVDTLTRAGMDFMTH
Corvol_Ether_B       ---------------------F-DFPWPSA-C-HPHARQAEQGALAFAER
Delta-cadinene       --------------------GSA-IWDVLSP-H-SPHAAAADGKTLVWVEA
Hedycaryol           ----------------IDGVRL-VLPFPSE-V-SPDVDRARERHYGWVAQ
Intermedeol          ------------------QPAF-HMPFTAR-T-NPHLERTRRWLRQWAHR
Linalool             ---------------------F-AVPAPSR-V-SPDLARARARHLDWVHA
Odyverdiene_A        -------------------PDF-TTPFRYR-L-NPHLAEVTPRARQWMLE
Pimaradiene          ------------------------------------LKVLADLRSWAAE
Tau-muurolol         ------------------GAQI-VYPFTGA-I-SPYAGDVDQATLIWAEA
Terpentetriene       ---------------------EFEHEGR-R-NPNSAEAESAYSSIIAA
_-_-Isohirsut-4-ene  ---------------DPEAPALDELPFPRR-I-SPDFRAAHDRHLAWPRS

2-Methylisoborneol   CGFDEDE-A----------QKIRKV---RFGRLVMLAHPDCD--DPSRLL
4-epi-cubebol        ----DAE----------TVERYRQA---KYGWLSARTYPYAE--HH-TLR
African-1-ene        FGLGPSE-A--------SHQRLSQL---RLEVITARYYPFVE--PS-ALP
Alpha-amorphene      WGLTDDP-A--------VAAMISCT---RPAELAAFNGPDMD--SG-LLQ
Cembrene_C           HGFCNTE-A--------DRLVVANI---DAGAIVARWYPNPDFPVD-RLQ
Corvol_Ether_B       HGLVPTA-A--------YRSRLERT---RYGWLAARCYPDAD--DV-LLQ
Delta-cadinene       GELCGHDTA--------ESANLARI---RPGLLAAFCHPKAT--ED-DLT
Hedycaryol           HGLWPDR-K--------SEYAYKHA---DFPLFIAHVYPWAS--GE-DLD
Intermedeol          IGLLDPE-YTTPWPEQWSERKFEEA---DFALWTAMTHPDID--AD-ELN
Linalool             MDLVRGE-E--------ARRRYEFS---CVADIGAYGYPHAT--GA-DLD
Odyverdiene_A        SDLVDDT----------HMLEYEMA---RIPELMAAAYPGAS--AD-DLL
Pimaradiene          YPQVLEA----------------TPIEALAISTAAISPWRG--AN-ELR
Tau-muurolol         SGLLTDG----------LRQKSQRL---QYGVLAARAYPQAD--RD-TLQ
Terpentetriene       LDLQESD-Y--------A-VISGHS---RIVGAAALVYPDAD--AE-TLL
_-_-Isohirsut-4-ene  FGFLSTE-A--------ERAHHLKG---QFPLIAAMFYPNAT--GS-ELD
                                                          *        *
                                  Asp-rich
2-Methylisoborneol   IGAKLN-MAWWAADDYYADDSE-----LGAD--PKLLPPRLLLA-MTA--
4-epi-cubebol        LVSDWC-VWLFAFDDAF-CESD-RR-AAE------IA-RALPQL-YAV-L
African-1-ene        LLALHM-AWAWCVDEQF-DDGPAGR-DPR------WCLTALRGL-RDAVL
Alpha-amorphene      IAANQI-AYQFVFDDRA-EDIG-RH-SPG-----RLL-PMLSES-VAI--
Cembrene_C           MVTDFL-YLYFLIDDLR-FEVI-NS-DTGLAGPIALF-AQHLDL-WEY--
Corvol_Ether_B       LCADYF-IWFFIVDDLF-VDRV-DT-LSE-----RTI-PNLTAM-IDV-L
Delta-cadinene       LITKWM-AWLFLLDDRI-DESDLGR-DAD------LLDGHLQDL--QGVA
Hedycaryol           LVTDCV-GWAWLWDDSL-DRQ--AR-FP------WTE-DVLEAYFYGM-T
Intermedeol          LVTGWH-VALWFVDDLA-LPLF-GR-IDDRAAAQRQV-DRLLEF-LPV-D
Linalool             LCVDVL-GWTFLFDDQF-DAGD-GR-ERD------AL-AVCAEL-TDLLW
Odyverdiene_A        LSCDLM-GIMFAIEDED-CGSH-PRHSVA------GIATRCKAM-IQV-M
Pimaradiene          LSAPDVRCGPTPLDDHV-EQNV--R-SLD------EL-DDLFGR-CEA--
Tau-muurolol         IAADWI-AWLFFMDDQC-DEAGIGR-DPQ------RLAALHERF-LAV-L
Terpentetriene       AASLWT-ACLIVNDDRW-DYVQ--E-DGGRLAPGEWF-DGVTEV-VDT--
_-_-Isohirsut-4-ene  IGVDQQ-SWYFLFDDAL-DEQW-GG-SPE------RV-RHLVGL-V---Q
                                  ::
```

```
2-Methylisoborneol    --MDPPPPAGEFTPPLEAAIAEERVLVALGRGIDYLS-----QYATPEQV
4-epi-cubebol         EDLDVG--SE-VDDVFAKSLLEIKG---------RIA-----AYGDDEQL
African-1-ene         G-G--S--HI-PGNPLEAAAADFHR---------RLG-----EVHSPRWV
Alpha-amorphene       --LRDG--QP-PTTPLGAALADLHR---------QVQ-----ERCTPAQA
Cembrene_C            --PQAH--RREELDLFHQAIHDLAS---------RMA-----ELTTPTKA
Corvol_Ether_B        DHHRPG-A---EPVFGEHAWLDVCT---------RLR-----AYLSDEHF
Delta-cadinene        L-G--I--RT-ASGPMSRALEEIIT---------QAS-----AGMGDAWQ
Hedycaryol            DPSREP-AEA-VAVPLVHAWRTLNK---------RLH-----ARTSTAWR
Intermedeol           ALPRPL--LV-PRNPVERAFAELWP---------RTA-----PSMTPVWR
Linalool              K-GTAA--TA-ASPPIVVAFSDCWE---------RMR-----AGMSDAWR
Odyverdiene_A         GG-VDP---G-ADDPVVLAFSDTWH---------RLC-----DGMSDTWV
Pimaradiene           -IVRGG--DRDDGHPLLASLSGWQS---------ALERAPHYPKLAGLWG
Tau-muurolol          EG-DPP--DS-GDWNLTRALADIRR---------RLA-----ARATDDWL
Terpentetriene        --WRTA--GPRLPDPFFELVRTTMS---------RLD-----AALGAEAA
_-_-Isohirsut-4-ene   EG-VTG--QA-SPLPPAAAFADMRR---------RSC-----HGMPEDWI
```

PP Sensor

```
2-Methylisoborneol    QRTCYATFS----MFVSW-GTYAAWR-------YTGDYPPAWKYLAARQH
4-epi-cubebol         DRWRNVTKD----YLFAQ-VWEAANR-------EDEVVPSLEDYIFMRRR
African-1-ene         RDYTDVTTR----WLWSY-YAESLDR-------ATDRHQPLPDYRRHRQV
Alpha-amorphene       ARWAWNSRE----YVHGL-LYEAVAQ-------AHPAPVESGLCRSIRSL
Cembrene_C            ARMRRSING----WFLAL-LREIALF-------NDDHAVMAEEYLPIRVV
Corvol_Ether_B        QRFAHGMRM----WAATA-GLQIANH-------LGADTVDVAPYETIRRH
Delta-cadinene        LRFRRNISD----YLLAC-VWQAAHR-------QAGEFPDPEVFPHWRRA
Hedycaryol            ARHEAHWRA----TFKGY-LQEARNN-------ATETIPTLEEYFDLRRK
Intermedeol           LRFRGDVER----FLRGV-LREIDRPAHENGEGAGGRAADPIEYVQDRRE
Linalool              RRTVHEWVD----YLAGW-PTKLADR-------AHGAVLDPAAHLRARHR
Odyverdiene_A         VRHRSSWKD----FLDNHHTWEPVVV-------EKRGMPTLEDYLWERAY
Pimaradiene           DRFAEALRGERYDWTAG----LARDR-------G-EGPSDPQEYLTYAAS
Tau-muurolol          RRFGEHVRL----YFTAN-RWEAANR-------CQSIVPNVATYCAARLF
Terpentetriene        DEIGHEIKR----AITAM-KWEGVWN-------EYTKKTSLATYLSFRRG
_-_-Isohirsut-4-ene   RRSAAHWSS----YLDHH-VHEARSR-------QSGAPMPLSTYLRVRRH
```

NSE Triad

```
2-Methylisoborneol    DS-FYTSMTLIDPI--GGYVLPPDLFFDPRVRHAAFLAGTAVVLVNDLLS
4-epi-cubebol         TGAMLTVFALIDVA--SGRSLSADEWRHPGMRAITESANDVVVWDNDLIS
African-1-ene         TSGEHMFFMLAEHG--AGRELPDSVSRLPAYIALQDAAAEHMGLLNDIVS
Alpha-amorphene       IAGVEPFYPLCEAA--QRCELAPEELHHPAMRRLSRLSADAAVWIPDLFS
Cembrene_C            TVASRLMIDVNGFI--CPAEVPGDEWYSLKVQAAAEAAMSVCLYDNELYS
Corvol_Ether_B        TSGTNPCLALADAA--KHGPVTPAEYHSPPVQRLVLHANNVVCWSNDVQS
Delta-cadinene        FGAIMPSFDLIERT--DGGALPSCVYYSRPYQSLLTAAADLVCWTNDLMT
Hedycaryol            TSGPETTFDWIEAA--GRFEVPTAIHATEAMLRLRRDAVDLISISNDLVS
Intermedeol           FGGLPMTSTLMEH---GQQEIPEHIYRLRSFQSALRAFADVISLHNDIVS
Linalool              TICCRPLFALAERV--GGYEVPRRAWHSSRLDGMRFTTSDAVIGMNELHS
Odyverdiene_A         SSGMYVLYDWSERFSADRAEIPQPALEDPRLATLHRNCIYTIIAINDTHS
Pimaradiene           SNAWITHFP---RW--ATSDRDDLLDGLPVLDNALEAIEVAVRLSNDLAT
Tau-muurolol          SGAVYACFDLIELA--AGIDLPFYARYHAAVQQLERTANNIICWCNDMLS
Terpentetriene        YCTMDVQVVLDKWI--NGGRSFAALRDDPVRRAIDDVVVRFGCLSNDYYS
_-_-Isohirsut-4-ene   TIGVAPVIDLAERL--SSCVLPDHLYALPHLSVMREMTKTFIICDNDIVS
                                                                  :  :

2-Methylisoborneol    VAKDLADE--------QPP-VNMVLQIAADRGCSIQEATEITVKFHNDLV
4-epi-cubebol         YAKESNSG-------NS-R-NNLVNVLAEHRHYSRQEAMEEIGEMRNQAI
African-1-ene         LPKEEPIG--------D-V-HNAVTLVSRHSRVPAAEAAEAVNDMLTDCV
Alpha-amorphene       AVKEQRAG--------G-M-INLALAYRRTHRCSLPAAVTLAVRHINSTI
Cembrene_C            AGKEQWLKSRATAHDRR-P-RNLVALIQAQTGGSTEHALQEVAEYRNRTV
Corvol_Ether_B        LKMELNQP--------GQY-WNMAA-IYAHRGLSLQQAVDLVALRVRGEI
Delta-cadinene        VDKEAAHG--------D-L-HNLVLVTEHDRHQDRRTASAAVSAACEQRM
Hedycaryol            ARNEWSEG--------N-T-DNIVIVLAYQEQCTWPEAARMAESIAHSIV
Intermedeol           YEREIAEG--------T-VDNNGVEVLRRALGCDLQQATTALNAMLTGRV
Linalool              FEKDRAQG--------H---ANLVLSLVHHGGLTGPEAVTRVCDLVQGSI
Odyverdiene_A         LEREIRRN-------DP-V-PNLLKVLMHHERLTVEQSVERAKKMLADAI
Pimaradiene           FERERAEP--------G-Q-NNILMYDTSP------D-------WVHDEL
Tau-muurolol          YPKEMQHG-------DV---HNLVLAIRQEYQCSLSEALQQALLLHDRET
Terpentetriene        WGREKKAV--------D-K-SNAVRILMDHAGYDESTALAHVRDDCVQAI
_-_-Isohirsut-4-ene   LEKDAALG--------E-Q-NNLVLCLERDHGLSRPQAVDRALRRRAEAL
                         :                *
```

```
2-Methylisoborneol   HDFR-ERH-QKLQAV---------PNVELQRF-LRGLRGWMGGAF--EWH
4-epi-cubebol        ADMV-AVR-PSLEA------L---GSDAVLAY-VRGLEFWISGSV--DYS
African-1-ene        HRML-AAE-DELPGQLDAVGVTERTRADALAM-VEAIKAHTRGNF--DWH
Alpha-amorphene      REFE-DLY-GEVRPE---------LSPSGIGY-VEGMAGWIRGCY--FWS
Cembrene_C           CLYL-NLR-SQLEKT---------ASPALLAY-LSVLDGVISGNL--DAH
Corvol_Ether_B       ASFQ-SLA-LTL---------EPHASRPLRGF-VDGLRHWMRGYQ--DWV
Delta-cadinene       RAHT-SAR-RDLTGLTAALGLPDTVRTHADDC-AASLLVWVRGHL--EWG
Hedycaryol           ENFL-ATE-QQLLASDLYQALAQDERADTDRF-INCIKHWIGGSH--TWH
Intermedeol          DTLR-HMVGTELPRVLGDEHLPERLAARITGY-LETLTTATAGSY--AWH
Linalool             ESFL-RLR-SGLPELGRALGV---EGAVLDRY-ADALSAFCRGYH--DWG
Odyverdiene_A        EEYL-EVE-FEYLAHWWRSGLPPRQMGAVEQR-LRDMRNWTSSNC--RWH
Pimaradiene          DRHSRKAQ-EQLDPL-ATAGFP--PAVELLRL-LDWSVTFYSGADFRGWG
Tau-muurolol         DTFM-KTQ-KQLPR------FNPAVDMALERY-IDGLQYWICANR--DWS
Terpentetriene       TDLD-CIE-ESIKRS---GHLGS-HAQELLDY-LACHRPLIYAAA--TWP
_-_-Isohirsut-4-ene  ELFT-GAH-RALLGSTATGHLDPAERDLLQRYCTEALQTTIRGAY--DWH
                                                                    .
                          RY
2-Methylisoborneol   N-SNPRY-KNSNGAS------Q-PDS---------------------QSL
4-epi-cubebol        L-TSSRY-TDAWRTARQ----P--SI-------------------------
African-1-ene        F-EVSRY-AGPGQLDEAG---S-PAYVA-DVLGG-----ART-----GAD
Alpha-amorphene      R-TVPRY-ADTL--T------A-PAG-------------------------
Cembrene_C           A-TSSRY-HNPDG-H------H-PHAIAFTPLRTTDECSARAHTPIAPPI
Corvol_Ether_B       ENDTLRY-ADAFIAEDA----D-DTAV-----R------------------
Delta-cadinene       L-ETPRY-RPGTTGTGT-----------------------------------D
Hedycaryol           L-DCPRY-KVPATRPLQ----V-PPKSTSQPSRH-----PEV-----SGI
Intermedeol          D-LTGRF-RQSAAPARTLVPHG-PTGFGTSAARL-----ART-----ASG
Linalool             R-GASRY-TTRDHPGDLG---L-ENLVA------------RS-----S-G
Odyverdiene_A        C-IVPRY-DHVARDPER----DRPGLPT-QPAG-------R------QGK
Pimaradiene          ---SDRDLTGPS--G------L-PSDM-------------------------
Tau-muurolol         L-TAMRY-ALSGTEPVL----R--TR----FA--------------------
Terpentetriene       T-ETNRY-R-------------------------------------------
_-_-Isohirsut-4-ene  H-ASTRY-HG-----------------------------------------
                          *

2-Methylisoborneol   A--------
4-epi-cubebol        -R-------
African-1-ene        ---------
Alpha-amorphene      --------L
Cembrene_C           AWWWEQLDQ
Corvol_Ether_B       T--------
Delta-cadinene       ---------
Hedycaryol           Q--------
Intermedeol          ---------
Linalool             ---------
Odyverdiene_A        DT-------
Pimaradiene          ---------
Tau-muurolol         HS-------
Terpentetriene       ---------
_-_-Isohirsut-4-ene  ---------
```

Appendix Figure 3: Multiple sequence alignment of the amino acid sequences of the class I TPSs used in the class I HMM. The Asp-rich, pyrophosphate (PP) sensor, NSE triad and RY motifs are labelled in red. The gene names are shown in Table 3.1 in chapter 3.2.1.1.

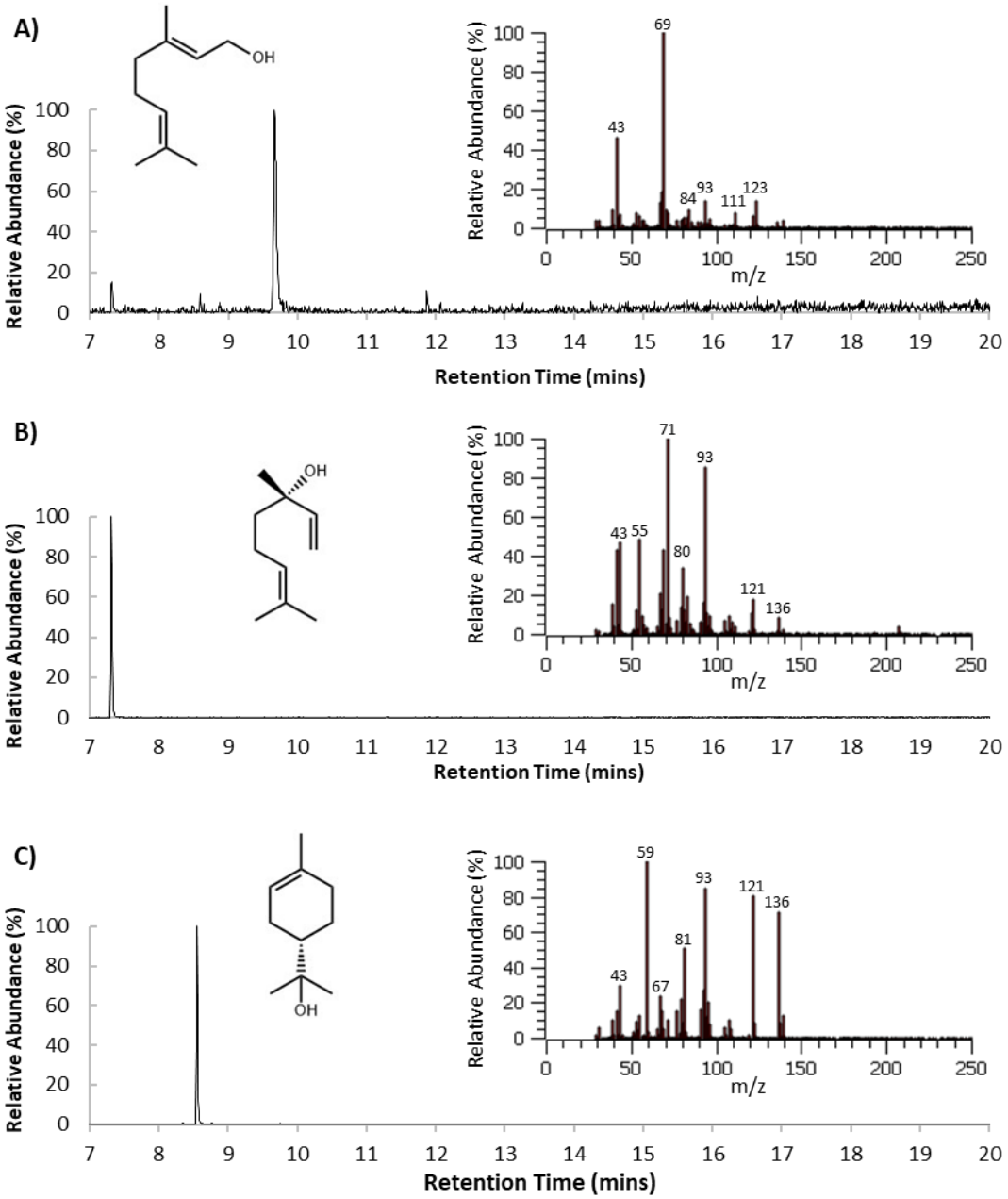
8.5.2 Class II terpene synthase sequence alignment

```
Blr2149         MRNGWRNSYSGPTRICSNTRHAHERARPHRVAPVPLRSAGGRVHPALCPS
Cyc1_KitGr      MKDRA-----------------------------------------------
Haur_2145       MSLIV-----------------------------------------------
Rv3377c         MET-------------------------------------------------
Sare_1288       MSADLGA-----------------SAPAAVEPVPA----------------
                *

Blr2149         PRRRSQRLPAAAHVSANGTSRVNALSEHILSELRRLLSEM--SDGGSVGP
Cyc1_KitGr      ------------------------------------A--DPVTKFSP
Haur_2145       -------------------------DILIDDLRALIRDLG-QNGGLMSP
Rv3377c         ---------------------------------FRTLLAKA--ALGNGISS
Sare_1288       --------------------------GRADTAESLVAELIRVPAGQVSP
                                                                ...
                                                  QW Motif 1
Blr2149         SVYDTAQALRFHGNVTG-R-QDAYAWLIAQQQADGGWGSAD-FPL---FR
Cyc1_KitGr      SPYETGQFLRISERADV-G-TPQIDYLLATQRPDGLWGSVG-FEL-----
Haur_2145       SVYDTSQALRLYPTPSEEHVWPAVNWLISQQQSDGGWGNPS-MPL---SR
Rv3377c         TAYDTAWVAKLGQLDDE-LSDLALNWLCERQLPDGSWGAE--FPFCYEDR
Sare_1288       SLYETARLVSLAPWLTG-H-AERVRHLLTSQRPDGGWGPPEGYAL-----
                : *:*.    :               *   * .** **      :

Blr2149         HAPTWAALLALQRAD-------PLPGAADAVQTATRFLQRQ-PDP-YAHA
Cyc1_KitGr      -VPTLGAVAGLSSRPEYA----DRAGVTDAVARACEKLWELALGEGGLPK
Haur_2145       AVPTLAAILALRRHC-------QRRSTFDGLLEAKRFLRRQ-LEY-WEKP
Rv3377c         LLSTLAAMISLTSNKHRR----RRAAQVEKGLLALKNLTSGAFEG-PQLD
Sare_1288       -VPTVSATEALLAELRTAPAAEPLIRATDAGLTVLTRWLSA------PRS
                  .* .*  .*                 :    .

Blr2149         VPEDAPIGAELILPQFCGEAASLL----------GG----V-AFPRHPA-
Cyc1_KitGr      LP--DTVASEIIVPSLIDLLSEVL-Q-RHRPAV-GGKAGQEQEFP-SPP-
Haur_2145       LPDNLPVGMELLLPYMLEEAYREEHQ-DDIDDV-P--IK--LRLN-IPLA
Rv3377c         IK-DATVGFELIAPTLMAEAARLGLAICHEESILGE--------------
Sare_1288       LP--DTPAIDLIVPALAAAINRHLVE-ADLPSALGH-WRAAARLR-LPA-
                :    . . ::: * :

Blr2149         ----LLPLRQACL--VKLGAVAMLPS-GHPLLHSWEAWGTSP-----TTA
Cyc1_KitGr      ----GANAELWRQLSDRIARGQAIPK---TAWHTLEAFHPLPK-QFAATV
Haur_2145       PYRELIALGEHKR--SLIQQ--KKPRAGTAPVYSWEAWASHA-----DPE
Rv3377c         ----LVGVREQKL--RKLGG-SKINK-HITAAFSVELAGQDGVGMLDVDN
Sare_1288       ----GMDDRRLAAVHGLIGAGAALPE---KLLHALEVVGSAAH--GVRGV
                              :             .: *

Blr2149         CPDDDGSIGISPAATAAWRAQAVTR-GSTPQVGRADAYLQMASRATRSGI
Cyc1_KitGr      TPAADGAVTCSPSSTAAWLSAVGTD----AG-ASTRAYLDEAQSR----Y
Haur_2145       LIDGSGGIGHSPAATAAWLFAANHNPNLRNEIAGAENYLRQASLATSESA
Rv3377c         LQETNGSVKYSPSASAYFALHV-K-----PGDKRALAYISSIIQA----G
Sare_1288       RPTRSGIVGASPAATAAWLGSPAGG----HRHPGASAYLERVVRQ----H
                    .* :  **:::* :               :  *:

Blr2149         EGVFPNVWPINVFEPCWSLYTLHLAGLFAHPALAEAVRVIVAQLD---AR
Cyc1_KitGr      GGAIPMGSSMPYFEVLWVLNLVLKYF-PDVPIPREI----IEEIA---AG
Haur_2145       PCIMPTAWPIPRFEQSFSLYALVTGGILDFPSIQDVLKPQIADLH---QA
Rv3377c         DGGAPAFYQAEIFEIVWSLWNLSRTD-IDLSD-PEIVRTYLPYLDHVEQH
Sare_1288       DALAPCATPITVFERAWVVATLARAG-LAVTQAPDLIPGLIAD-------
                    *       **  : :  :        .   :     :
                                  Asp-rich
Blr2149         -LGVHGLGP-ALHFAADADDTAVALCVLHLAGRDPAVDALRHFEIGELFV
Cyc1_KitGr      -FSDSGIGG-GPGLPPDGDDTAYANLAGDKLGAPTHPEILMKFWAEDHFV
Haur_2145       -LKPRGIGF-SDDFMPDGDDTAAAVAVLIAAGYPVDLAILNQFEREPYFV
Rv3377c         WVRGRGVGWTGNSTLEDCDTTSVAYDVLSKFGRSPDIGAVLQFEDADWFR
Sare_1288       -LTSVGTCA-GPGLPPDADTTAVTLYALAHLGFSVDLECLWRYETPDGFC
                 .   *    .     * * *: :  .    *       : ::     *

Blr2149         TFPGERNASVSTNIHALHALRLL-----GK---------PAA------GA
Cyc1_KitGr      SYPGEQTPSETVNAHALEYLNHL-----RMRRGITEFGAVED------AC
Haur_2145       AYHGELQPSISLTARAVHALDLA-----GV---------DIS------RW
Rv3377c         TYFHEVGPSISTNVHVLGALKQA-----GY---------DKCHP-RVRKV
Sare_1288       TWPGEDGFSLSTNAHVLDVVGLILTTDPGA---------DRRHVTAARRL
                ::  *   * : . :.:  :
```

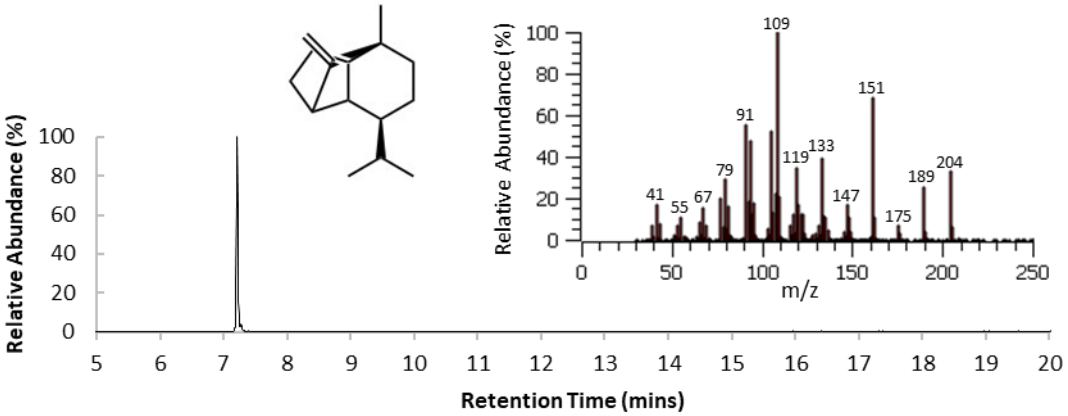
```
Blr2149         SAYVEANRNPH-GLWDNEKWHVSWLYPTAHAVAALAQ----GKPQ-WRDE
Cyc1_KitGr      AEWVISQQTED-GCW-YDKWNVSPYYSTAACVEALLDARKQDEPQLDSLR
Haur_2145       WKIFIDAQKLD-GSWSGDKWNTSWLYTTCHVLIALKN----SPYK-TAMK
Rv3377c         LEFIRSSKEPGRFCW-RDKWHRSAYYTTAHLICAASN----YDDA--LCS
Sare_1288       ADALRQRQQAD-GSW-QDRWHASPYYATMCCALALAGFPG-PGTAVTSLA
                   .   :        *  ::*: *  *.*     *
                          QW Motif 2
Blr2149         RALAALLQAQRDDGGWGAGRG-STFEETAYALFALHVMDGSEEATGRRRI
Cyc1_KitGr      RAREWLLRHQTDSGGWGMAEP--SPEETAYAVLALDLFAS-RGGEGAEEC
Haur_2145       EAVAALQVHQHPDGGWGIINR-STTVETAYAVLALQNLRE-AG-LLDDDD
Rv3377c         DAIGWILNTQRPDGSWGFFDGQATAEETAYCIQALAHWQR-HS--GTS-L
Sare_1288       RAASWIVDTQRANGSWGRWKG--TVEETAYAVQVLATVG--RG--RPG-A
                 *   :   *  .*.**      :  ****.: .*

Blr2149         AQVVARALEWMLARHAAHGLPQTPLWIGKELYCPTRVVRVAELAGLWLAL
Cyc1_KitGr      AAAISRAKEFFTDESR----ENPPLWMGKDLYTPFRIVDVTVMCGRAVVG
Haur_2145       IHMLQRGYNWLCIHYRPFRMKEYQCWLNKEIYCPQRIDRAYELSAML-AV
Rv3377c         SAQISRAGGWLSQHCE---PPYAPLWIAKTLYCSATVVKAAILSALRLVD
Sare_1288       DEAIRRGRAYLTEGTTAH-DPGPPLWHDKDLYRPAMIVRAAVVAARHLAG
                   : *.  ::              *  * :* .  :  .  :..   .

Blr2149         RWGRRVLAEGAGAAP
Cyc1_KitGr      RY-------------
Haur_2145       TLGELKL--------
Rv3377c         ESN-----------Q
Sare_1288       AAGPATA--------
```

Appendix Figure 4: Multiple sequence alignment of the amino acid sequences of the class II TPSs used in the class II HMM. The Asp-rich and two QW motifs are labelled in red. The gene names are shown in Table 3.3 in chapter 3.2.1.2.

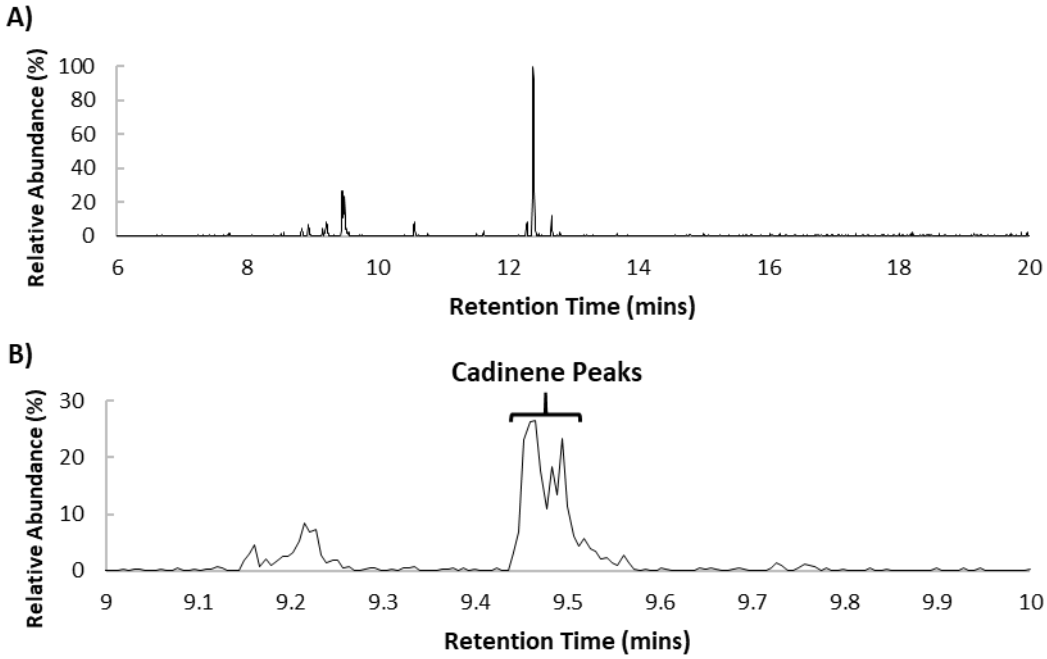
8.6 GCMS chromatograms and fragmentation patterns of authentic standards and samples



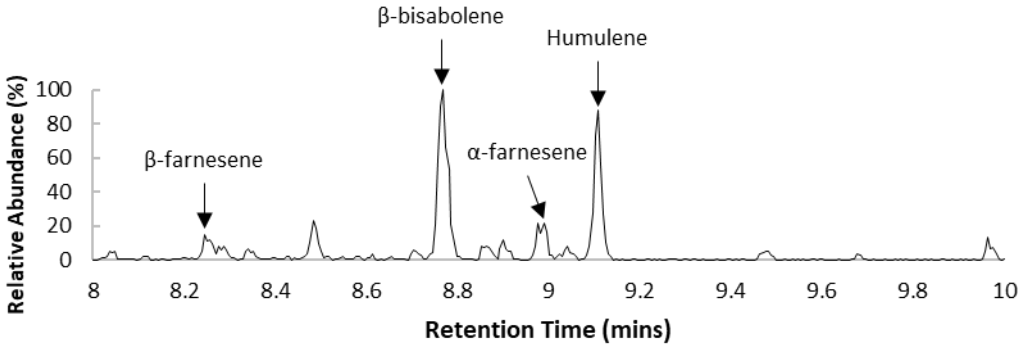
Appendix Figure 5: Authentic standard peaks for the monoterpenes, A) geraniol, B) linalool and C) α-terpineol on the GC-MS chromatogram. The retention times were 9.667 mins, 7.318 mins and 8.555 mins, respectively. The structures and fragmentation patterns are shown for each monoterpene.



Appendix Figure 6: GC-MS chromatogram, mass spectrum and the structure for the authentic standard of (+)-sativene (7.217 mins).

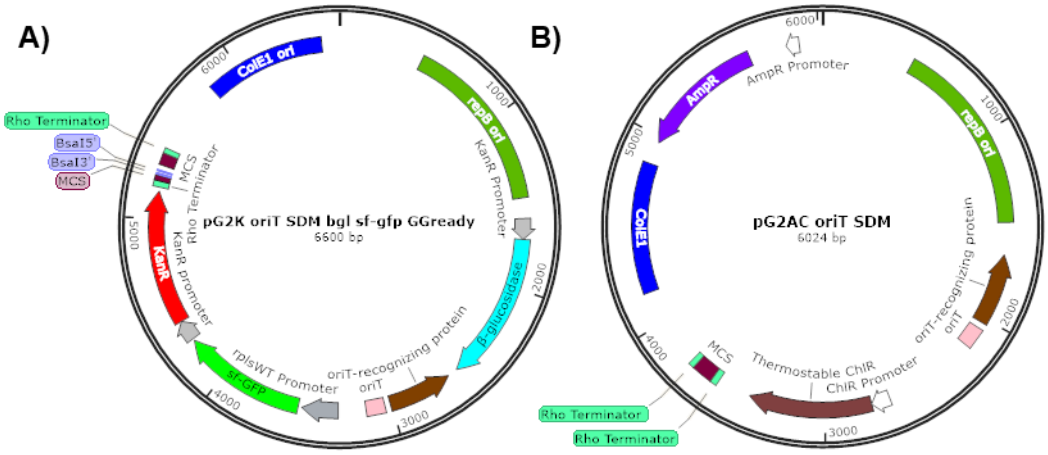


Appendix Figure 7: GC-MS cadinene peak from one of the L54A replicate assays at 37 ºC. A) The GC-MS chromatogram of the first replicate of the RoseRS_3509 mutant, L54A, at 37 ºC. B) The retention time of the three terpenes products labelled as “cadinene” between 9.4-9.6 mins.

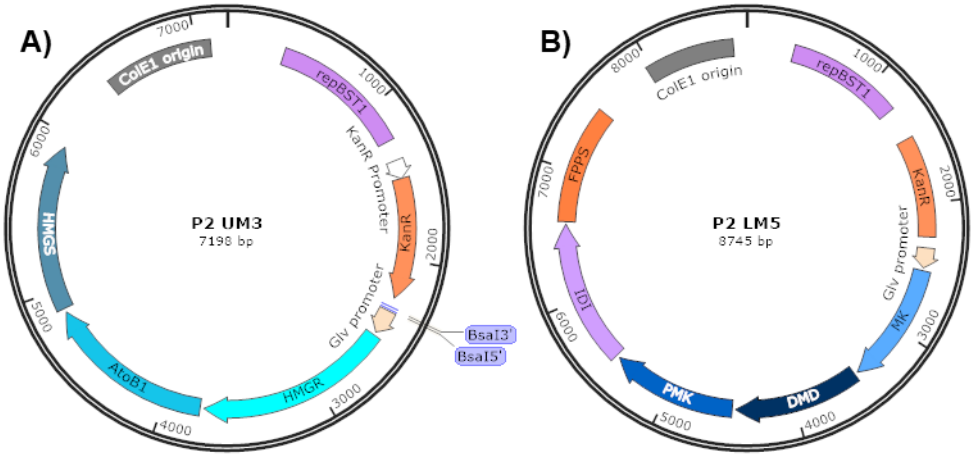


Appendix Figure 8: GC-MS chromatogram of the one in 100000 dilution of the farnesene, mixture of isomers sample (Sigma-Aldrich, UK). The retention times of the terpene products were β-farnesene (8.290 mins), β-bisabolene (9.769 mins), α-farnesene (8.902 mins) and humulene (9.109 mins).

8.7 *Parageobacillus thermoglucosidasius* plasmid maps

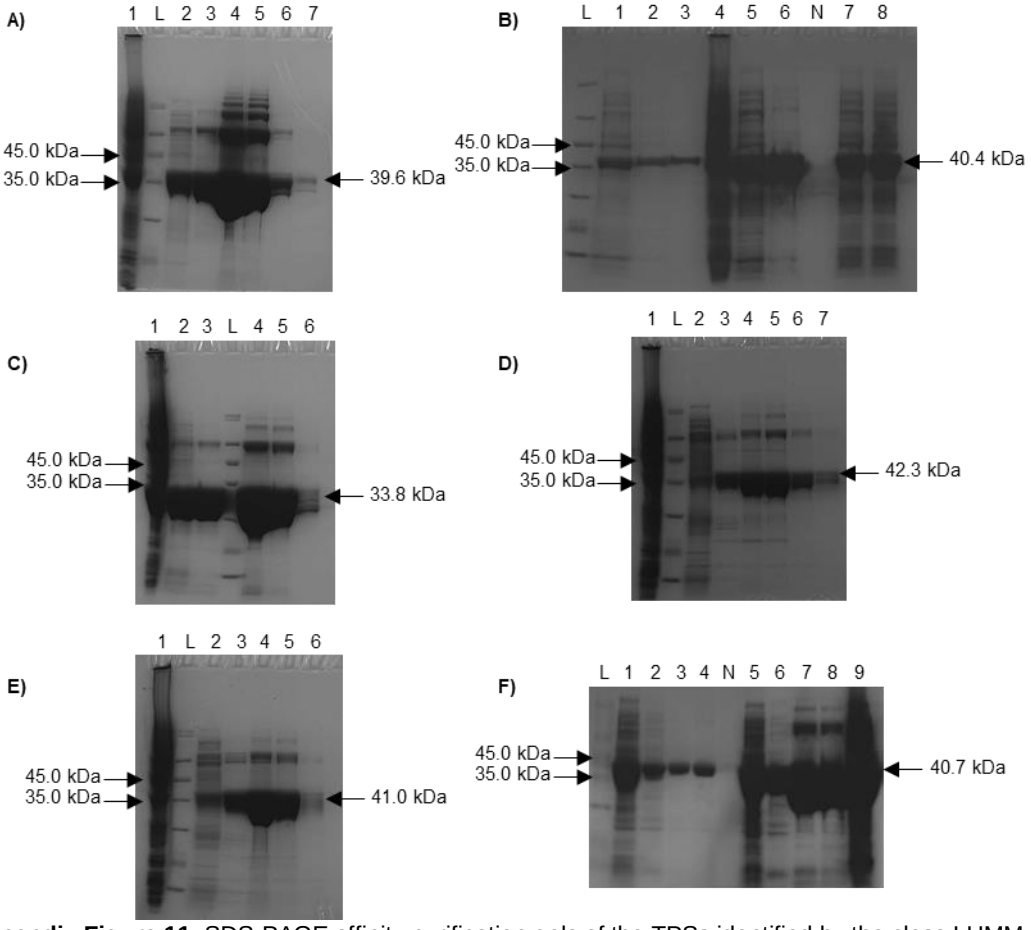


Appendix Figure 9: Plasmid maps of A) pG2K oriT SDM bgl sf-gfp GGready and B) pG2AC oriT SDM. To conjugate into *Parageobacillus* using the origin of transfer (oriT), *prpls* promoter and g*eoOpt_W296A* sequences were inserted into pG2K oriT SDM bgl sf-gfp GGready before inserting the promoter/gene construct into pG2AC oriT SDM by restriction with NheI and HindIII or SmaI and ligation. As the strain to be conjugated into, *P. thermoglucosidasius* cMev9, contained kanamycin resistance, pG2AC oriT SDM needed to be used as it had chloramphenicol resistance (c*hlR*).

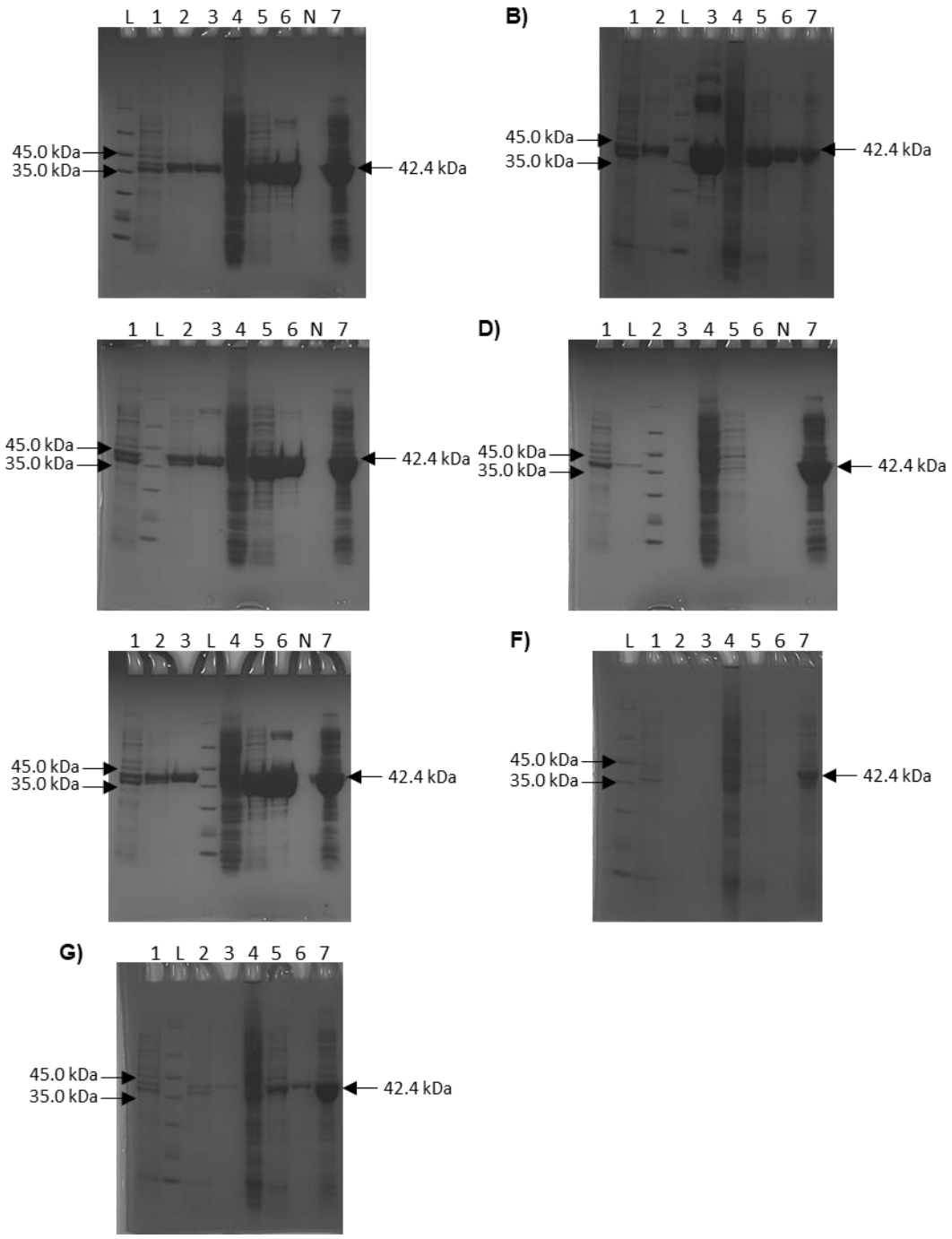


Appendix Figure 10: Plasmid maps for A) P2 UM3 and B) P2 LM5. These plasmids were designed by Dr Mathew Styles (University of Bath, UK) and contained the mevalonate pathway genes for forming pG2K oriT attP Mev9 plasmid shown in Figure 1.10.

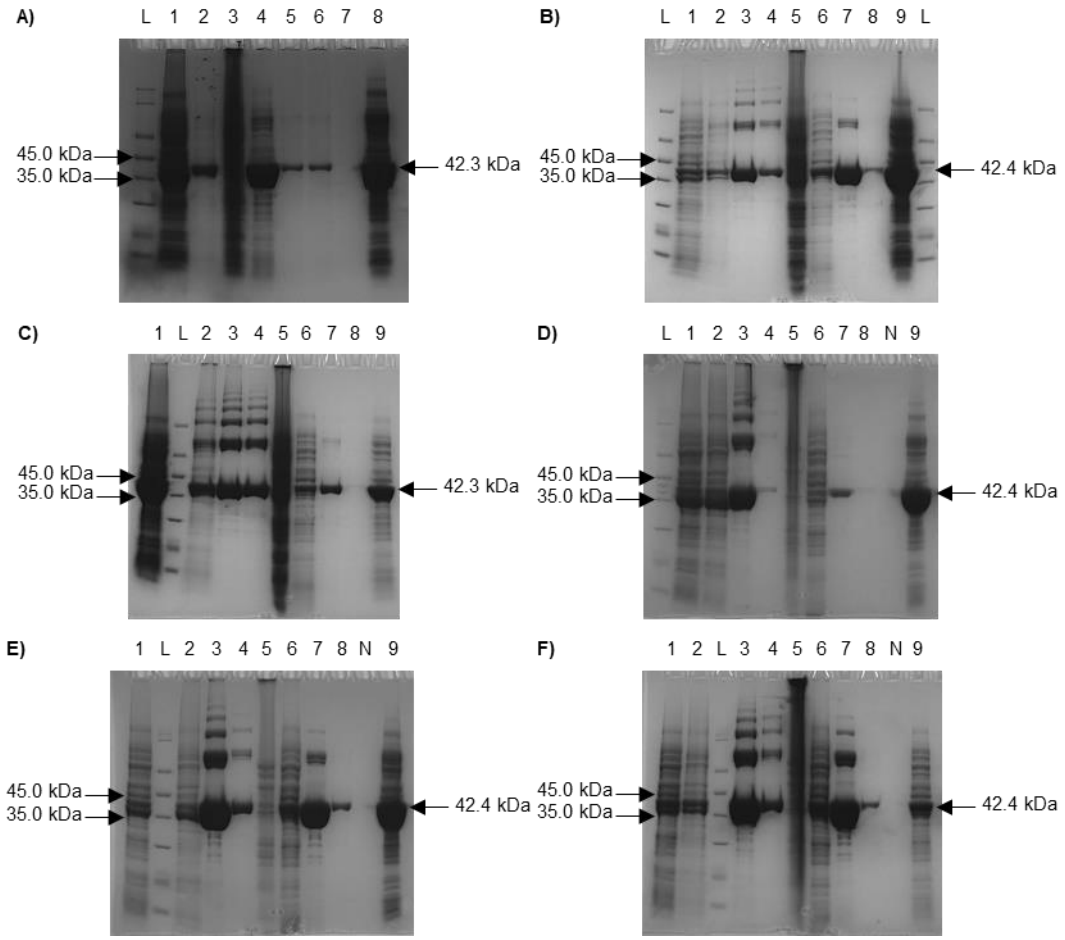
8.8 Affinity protein purification SDS-PAGE gels



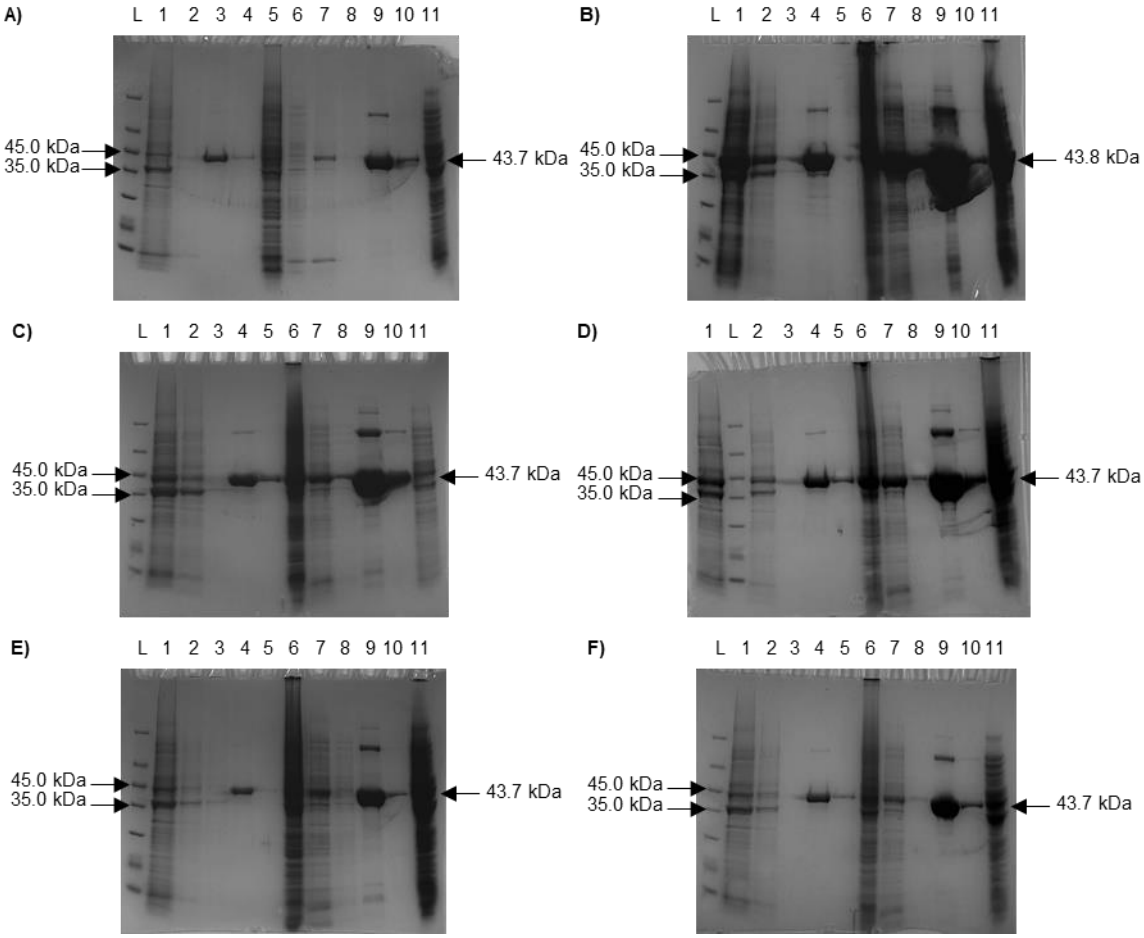
Appendix Figure 11: SDS-PAGE affinity purification gels of the TPSs identified by the class I HMM: A) Rxyl_0493, B) Tcur_3107, C) Tbis_3257, D) JKG1, E) FJSC11 and F) Tchrom JKG1 and Tchrom were purified and SDS-PAGE gels run by Carlos Amaya (Former MSc student, Leak laboratory) under my supervision. Lanes labelled L and N were Unstained Protein Molecular Weight Marker and nothing respectively. The lanes for A) Rxyl_0493, C) Tbis_3257 D) JKG1 and E) FJSC11 were Lane 1: soluble FT fraction, Lane 2: soluble buffer wash fraction 1, Lane 3: soluble buffer wash fraction, Lane 4: soluble elution fraction 1, Lane 5: soluble elution fraction 2, Lane 6: soluble elution fraction 3, Lane 7: soluble elution fraction. The lanes for F) Tchrom were: Lane 1, insoluble FT fraction, Lane 2, insoluble wash fraction, Lane 3, insoluble elution fraction 1, Lane 4, insoluble elution fraction 2, Lane 5, soluble FT fraction, Lane 6, soluble wash fraction, Lane 7, soluble elution fraction 1, Lane 8, soluble elution fraction 2, Lane 9, insoluble cell debris pellet. The lanes for B) Tcur_3107 were Lane 1: insoluble FT fraction, Lane 2: insoluble buffer wash fraction, Lane 3: insoluble elution fraction 1, Lane 4: soluble FT fraction, Lane 5: soluble buffer wash, Lane 6: soluble elution fraction, Lane 7: insoluble cell debris pellet 1, Lane 8: insoluble cell debris pellet 2.



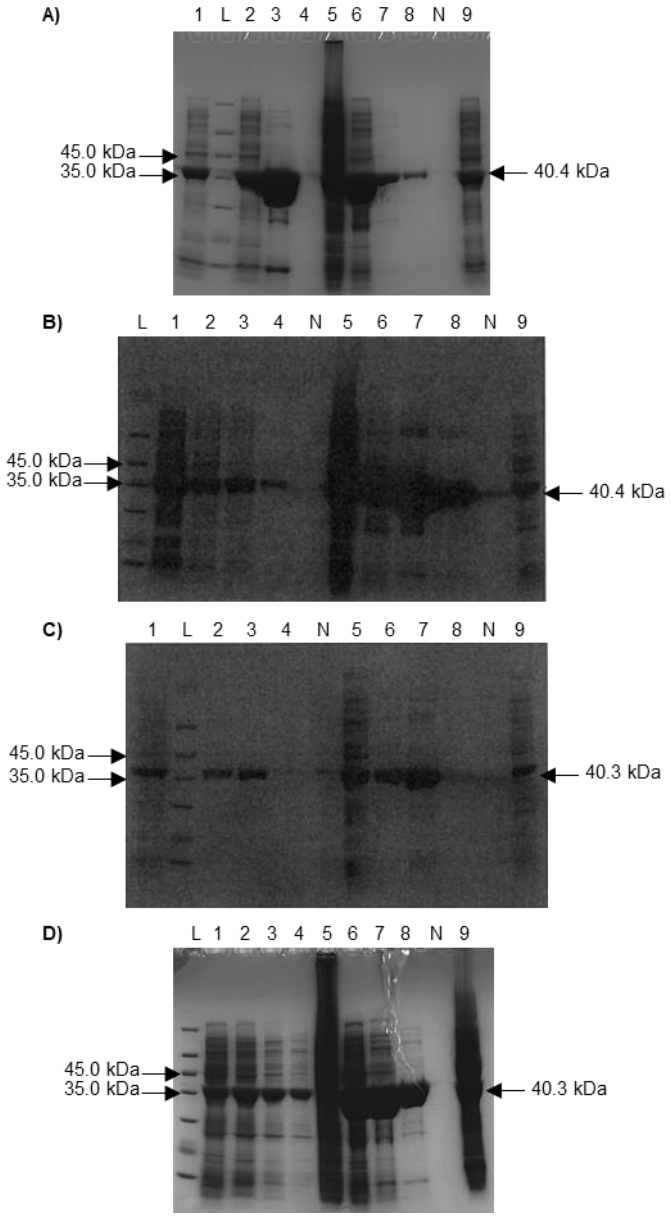
Appendix Figure 12: SDS-PAGE affinity purification gels of the SSCG_R6 mutants (42.4 kDa) determined using HoTMuSiC: A) G24I, B) P39I, C) S100Q, D) A115P, E) G209A, F) D224C and G) T293L. L – Unstained Protein Molecular Weight Marker. Lane 1: insoluble FT fraction, Lane 2: insoluble buffer wash fraction, Lane 3: insoluble elution fraction, Lane 4: soluble FT fraction, Lane 5: soluble buffer wash fraction, Lane 6: soluble elution fraction, Lane 7: insoluble cell debris pellet, N: Nothing.



Appendix Figure 13: SDS-PAGE gels of the fractions of the affinity purification of A) SSCG_R6 and its mutants, B) SSCG_R6_G46H, C) SSCG_R6_H106L, D) SSCG_R6_T207H, E) SSCG_R6_V274M and F) SSCG_R6_T311L. The size of all these proteins were 42.4 kDa except SSCG_R6 and SSCG_R6_H106L which was 42.3 kDa. The lanes labelled L and N were Unstained Protein Molecular Weight Marker and nothing, respectively. The contents for each lane for each gel except SSCG_R6 were 1: insoluble FT fraction, 2: insoluble buffer wash fraction, 3: insoluble elution fraction 1, 4: insoluble elution fraction 2, 5: soluble FT fraction, 6: soluble buffer wash fraction, 7: soluble elution fraction 1, 8: soluble elution fraction 2, 9: insoluble cell debris pellet. For SSCG_R6, the lanes contained: 1: insoluble FT fraction, 2: insoluble buffer wash fraction, 3: soluble FT fraction, 4: soluble buffer wash fraction, 5: soluble elution fraction 1, 6: soluble elution fraction 2, 7: insoluble cell debris pellet.

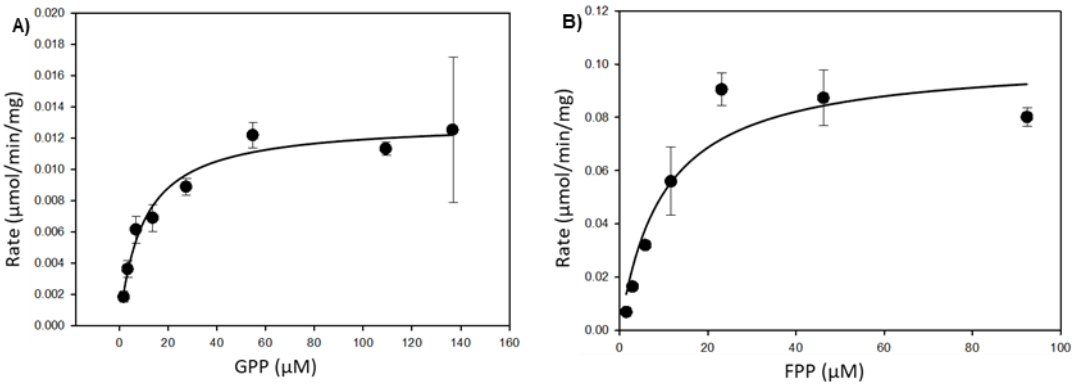


Appendix Figure 14: SDS-PAGE gels of the fractions of the affinity purification of A) SdS and its mutants, B) SdS_G45K, C) SdS_I106L, D) SdS_R206H, E) SdS_T273K and F) SdS_V309P. The size of all of the mutants were 43.7 kDa except SdS_G45K which was 43.8 kDa. The contents for each lane were L: Unstained Molecular Weight Marker, 1: insoluble FT fraction, 2: insoluble wash fraction, 3: insoluble buffer wash fraction, 4: insoluble elution fraction 1, 5: insoluble elution fraction 2, 6: soluble FT fraction, 7: soluble wash fraction, 8: soluble buffer wash fraction, 9: soluble elution fraction 1, 10: soluble elution fraction 2, 11: insoluble cell debris pellet.

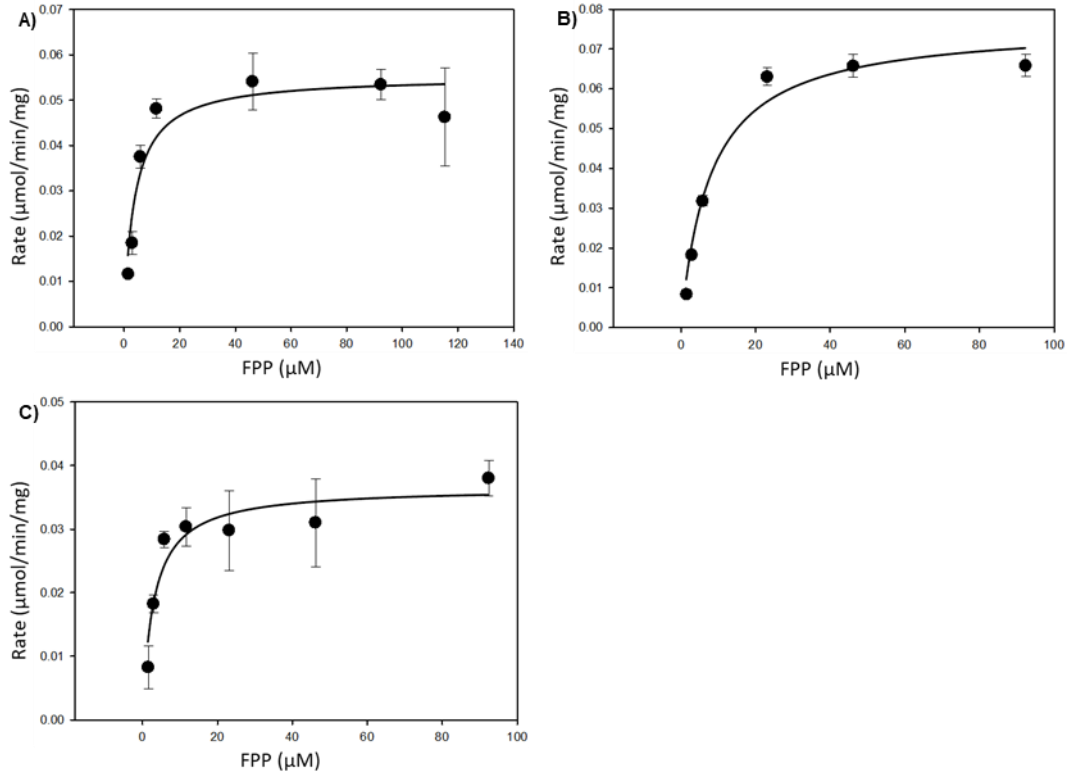


Appendix Figure 15: SDS-PAGE gels of the fractions of the affinity purification of A) RoseRS_3509 and its mutants, B) F78A, C) Y310A and D) W296A. F78A and Y310A were purified and analysed by SDS-PAGE gels by Charlotte Tidd (Former undergraduate student, Leak laboratory) under my supervision. The contents for each lane were L: Unstained Molecular Weight Marker, 1: insoluble FT fraction, 2: insoluble wash fraction, 3: insoluble elution fraction 1, 4: insoluble elution fraction 2, 5: soluble FT fraction, 6: soluble wash fraction, 7: soluble elution 1 fraction, 8: soluble elution fraction 2, 9: insoluble cell debris pellet, N: Nothing.

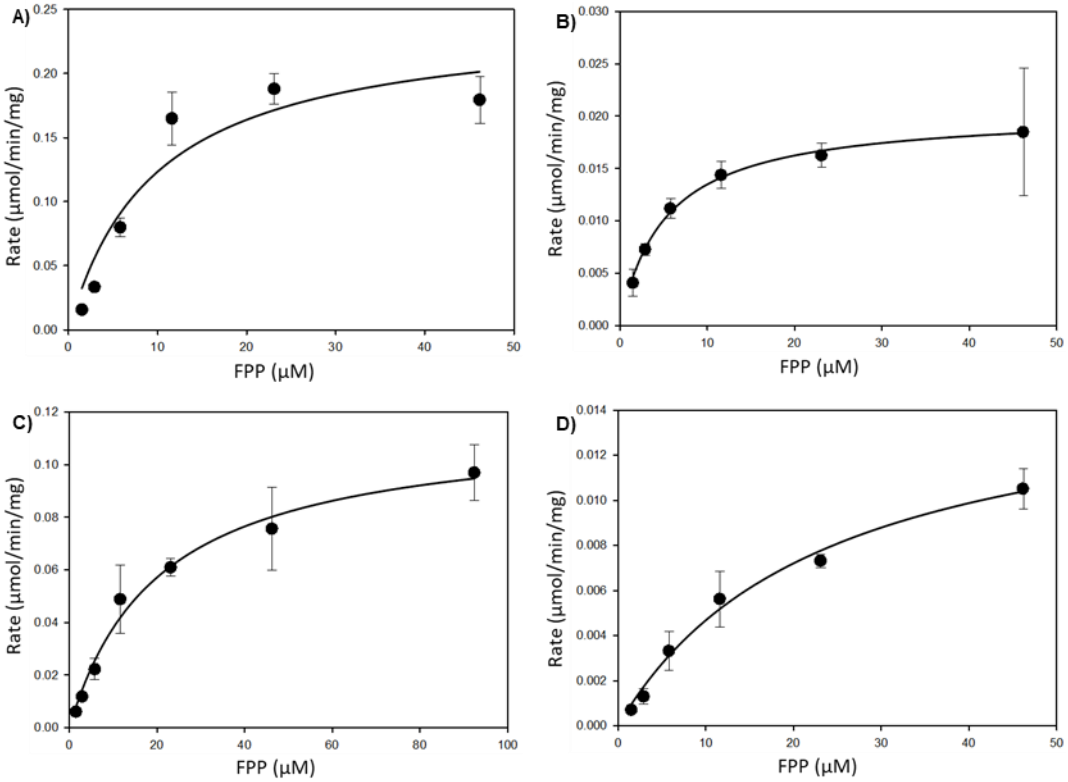
8.9 Michaelis-Menten Kinetic Curves



Appendix Figure 16: The effect of GPP (A) and FPP (B) concentration on the rate of Tchrom at 50 ^{o}C using the malachite green assay.



Appendix Figure 17: The effect of changing the concentration of FPP on the rate of the protein variants of SSCG R6: A) G24I, B) S100Q and C) G209A at 30 ^{o}C using the malachite green assay.



Appendix Figure 18: The effect of changing the FPP concentration on the rate of the protein variants of RoseRS_3509 (A), B) F78A, C) W296A and D) Y310A at 50 ^{o}C using the malachite green assay.